THE DUAL NATURE OF CAUSATION
TWO NECESSARY AND JOINTLY SUFFICIENT CONDITIONS

Caroline Torpe Touborg

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2018

The Dual Nature of Causation

Two Necessary and Jointly Sufficient Conditions

Caroline Torpe Touborg

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of PhD
at the University of St Andrews

20th April 2017

1. Candidate's declarations:

I, Caroline Torpe Touborg, hereby certify that this thesis, which is approximately 79,950 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2013 and as a candidate for the degree of PhD in September 2013; the higher study for which this is a record was carried out in the University of St Andrews between 2013 and 2017.

Date …………… signature of candidate ……………………………………

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …………… signature of supervisor ……………………………………

Date …………… signature of supervisor ……………………………………

3. Permission for publication: (*to be signed by both candidate and supervisor*)

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

PRINTED COPY
a) No embargo on print copy

ELECTRONIC COPY
a) No embargo on electronic copy

Date …………… signature of candidate …………………………………

Date …………… signature of supervisor …………………………………

Date …………… signature of supervisor …………………………………

Acknowledgements

First and foremost, I would like to thank my supervisors Professor Katherine Hawley and Dr Derek Ball for their support and generosity throughout the writing of this dissertation. In particular, I would like to thank Katherine for always encouraging me to see the bigger picture and look for ways to make my proposals simpler, and I would like to thank Derek for his challenging questions and counterexamples that have helped me to develop my ideas. In addition, I would like to thank Sara Bernstein, Sarah Broadie, Aaron Cotnoir, Matthew McGrath, Laurie Paul, and Gabriel Uzquiano for interesting and enjoyable discussions.

I would like to thank all my colleagues in Arché, especially Alex Yates – for many good laughs and great walks; Ryo Ito, Fenner Tanswell, and Alper Yavuz – for being the best possible friends and office mates (not to mention football champions); and Claire Field, Josh Habgood-Coote, Bruno Jacinto, Martin Lipman, Poppy Mankowitz, Mark Thakkar, Ravi Thakral, and Michael Traynor for many great conversations. In addition, I would like to thank the members of the Metaphysics Research Group, and especially Lorenzo Lazzarini for co-organising the group with me this year.

I would also like to thank Marc Cole, Kate Hazel Stanton, and Leonie Smith for all our philosophical discussions; Clémence Arhoul and Laila Atalla for being wonderfully supportive fellow lovers of chocolate; and my fellow residents in Deans Court.

I am grateful to audiences in St Andrews and at Brown University, and to the Philosophical Quarterly, the Janet T Anderson Trust, and the Royal Institute of Philosophy for their financial support.

Finally, I am grateful to my parents for everything.

My paper, 'Hasteners and Delayers: Why Rains Don't Cause Fires', which is based on ideas presented in Part IV, has been accepted for publication in *Philosophical Studies*.

Caroline T Touborg

St Andrews, April 2017

Abstract

In this dissertation, I propose a reductive account of causation. This account may be stated as follows:

> *Causation*: c is a cause of e within a possibility horizon $\mathscr{H}$ iff
>
> a) c is process-connected to e, and
>
> b) e security-depends on c within $\mathscr{H}$.

More precisely, my suggestion is that there are two kinds of causal relata: instantaneous events (defined in Chapter 4) and possibility horizons (defined in Chapter 5). Causation is a ternary relation between two actual instantaneous events – the cause c and the effect e – and a possibility horizon $\mathscr{H}$.

I argue that causation has a dual nature: on the one hand, a cause must be connected to its effect via a genuine process; on the other hand, a cause must make a difference to its effect.

The first condition – namely, the condition of *process-connection* (defined in Chapter 6) – captures the sense in which a cause must be connected to its effect via a genuine process. This condition allows my account to separate causation from mere correlation, distinguish genuine causes from preempted backups, and capture how a cause must be at the right level of detail relative to its effect (Chapter 7).

The second condition – namely, the condition of *security-dependence* (defined in Chapter 8) – captures the sense in which a cause must make a difference to its effect. This condition allows my account to yield intuitively correct verdicts on the counterexamples to the transitivity and intrinsicness of causation, resolve the problem of profligate omissions, accommodate structurally isomorphic but causally different cases, and handle contrastive causal claims (Chapter 9 and 10).

Finally, my proposed account of causation logically entails restricted versions of three important principles of causal reasoning concerning the sufficiency of counterfactual dependence for causation, and the transitivity and intrinsicness of causation (Chapter 11).

Contents

PART I

Introduction

1

Introduction and methodology

What is causation? And why should we care about what it is?

Let me begin by suggesting an answer to the second of these two questions. There are two main reasons why we should care about what causation is. The first reason is that causation plays a central role in our everyday lives. Everything we do, and everything that happens to us, involves causation: we act on our surroundings to bring things about or make things happen, and our surroundings act on us. Correspondingly, our understanding of causation plays a central role in our daily decisions: in practically every daily decision, we are using our understanding of how to bring about the specific outcomes we desire – even in such mundane activities as getting the kettle boiling, getting the toast nice and brown, etc. The fact that causation plays such a central role in our lives provides the first reason why we should care about what it is.

The second reason emerges once we notice that causation is closely connected with other things we care about, such as morality, perception, knowledge, and reference. According to consequentialist theories of morality, for example, the rightness of an action is determined by its consequences, i.e. by its *effects*. According to causal theories of perception and knowledge, we perceive and gain knowledge of the external world by coming into causal contact with it. According to causal theories of reference, the name 'Socrates' refers to Socrates because there is a causal chain from the original naming of Socrates to our use of the name 'Socrates' today, etc. Although one may of course question each of these theories, it is hard to deny that causation is *in some way* connected with morality, perception, knowledge, and reference.

Further connections include the connections between causation and explanation, prediction, manipulability and responsibility.[1] Because of these connections, gaining a better understanding of causation might in turn help us gain a better understanding of other areas that we care about.

What, then, is causation? In spite of the central role of causation, philosophers have not yet been able to arrive at an account of causation that is 'at once clean, precise, and widely agreed upon'.[2] Indeed, some have begun to doubt whether it is realistic to hope for such an account. Thus, after a detailed critical survey of the literature, Paul and Hall write:

> 'After surveying the literature in some depth, we conclude that, as yet, there is no reasonably successful reduction of the causal relation. And correspondingly, there is no reasonably successful conceptual analysis of a philosophical causal concept. No extant approach seems able to incorporate all of our desiderata for the causal relation, nor to capture the wide range of our causal judgments and applications of our causal concept. Barring a fundamental change in approach, the prospects of a relatively simple, elegant and intuitively attractive, unified theory of causation, whether ontological reduction or conceptual analysis, are dim.'[3]

However, I believe that such pessimism is unwarranted. During the past decades, the debate about how best to give an account of causation has been intense. And, as Paul and Hall themselves point out, this debate has yielded important insights:[4] although the debate has not yet produced a fully successful account of causation, it has sharpened our understanding of the challenges that such an account must overcome, and it has produced a wealth of cases that bring into focus surprising features of the causal relation.

---

[1] Schaffer (2003), p. 29.
[2] Paul and Hall (2013), p. 1.
[3] Paul and Hall (2013), p. 249.
[4] Paul and Hall (2013), p. 1.

My aim in this dissertation is to build on these contributions to propose a new account of causation. More precisely, my aim is to propose an account of what may be called *singular causation*, as opposed to *general causation*. The relation of singular causation holds, to a first approximation, between token events. Thus, the claim that 'the lightning strike caused the forest fire' – where we are talking about a particular lightning strike and a particular forest fire – is an example of a singular causal claim. The relation of general causation, on the other hand, holds between event types. Thus, the claim that 'lightning strikes cause forest fires' is an example of a general causal claim. For this reason, general causation is also called 'type level causation'.[5]

In choosing to focus on singular rather than general causation, I am following a general trend in the philosophical literature on causation. Many of the most influential contributions to the debate have focused on singular causation. In particular, singular causation is the target of David Lewis's two seminal papers, 'Causation' and 'Causation as Influence',[6] and indeed, of most work in the counterfactual tradition.[7] In this literature, singular causation is sometimes also called 'token causation', 'actual causation' or simply 'causation'.[8] Since I will focus exclusively on singular causation in the following, I will follow the practice of calling it simply 'causation'.

Stated briefly, then, my aim is to provide an account of singular causation. As anyone familiar with the literature will know, however, accounts of causation differ from each other not merely in their contents, but also in their more specific aims and associated criteria of success. For example, an account aiming to fully characterise causation in non-causal terms differs in its aims and associated criteria of success from an account aiming to deepen our understanding of causation, while relying on our pre-existing ability to identify

[5] Lewis (1986*d*) and (2004*a*). See also e.g. Hitchcock (2007), p. 497; Paul and Hall (2013), p. 7; Woodward (2003), p. 40.

[6] Lewis (1986*d*) and (2004*a*).

[7] Hitchcock (2007), p. 497.

[8] Blanchard and Schaffer (forthcoming), p. 3; Halpern and Hitchcock (2015), p. 413-14.

causal structures. And an account aiming to capture our *concept* of causation again differs in its aims and associated criteria of success from an account aiming to lay bare what it is, out in the world, that we are targeting when we make causal claims.

All of these projects provide valuable insights into the nature of causation and it is often possible, while pursuing one kind of project, to draw on insights from projects with other aims and criteria of success. In developing an account, however, it is crucial at all times to keep to a single vision of the aims and associated criteria of success that one is attempting to satisfy. Otherwise, it is far too easy to subtly shift one's position, depending on what is most convenient in dealing with the case at hand.[9] I therefore devote the remainder of this chapter to the task of giving a precise characterisation of the aims and associated criteria of success of my proposed account.

I begin by drawing up the question that I aim to answer, using considerations of eligibility and charity to use to pick out the causal relation with greater precision (section 1). Next, I set out the aims and criteria of success for the biconditional I propose in this dissertation (section 2).

1. Drawing up the question

Some assumptions are so fundamental that they underlie the very questions we ask. In the following, I will be making one such assumption: that reality has structure. Early versions of this idea can be found in the writings of Armstrong and Lewis.[10] However, the idea has been developed and defended most fully by Sider in *Writing the Book of the World*.[11] Sider here gives the following initial characterisation of structure:

---

[9] See e.g. Paul and Hall (2013), p. 26.

[10] Armstrong (1978*a*) and (1978*b*); Lewis (1983).

[11] Sider (2011).

> 'Discerning "structure" means discerning patterns. It means figuring out the right categories for describing the world. It means "carving reality at its joints", to paraphrase Plato. It means inquiring into how the world fundamentally is, as opposed to how we ordinarily speak or think of it.'[12]

The idea that reality has structure is easiest to appreciate when we consider properties and relations. According to the idea that reality has structure, some properties and relations do objectively better at 'carving reality at its joints' than others. For example, the property of *being an electron* does objectively better at carving reality at its joints than the property of *being an electron or a cow*; the relation of *exerting a force on* does objectively better than the relation of *being one metre away from or exerting a force on*, etc.[13]

Sider develops and generalises this notion of structure so that it can be applied much more widely.[14] For my purposes here, however, the details of these further developments do not matter. What matters is simply the basic assumption that the world has structure and, in particular, that some properties and relations do objectively better than others at 'carving reality at its joints'. I do not think this is an unreasonable assumption. However, I will not attempt to defend it here. Rather, I will simply hope that this is an assumption that my reader will share. With this assumption in place, I will now move on to the main task of this section: delineating the question that my proposed account of causation aims to answer.

When we are considering a term such as 'cause', there are two kinds of questions we may ask. We may ask *semantic* questions, i.e. questions about what the term *means*. For example, we may ask: what is the meaning of 'cause'? And we may ask *metasemantic* questions, i.e. questions about what *determines* the meaning of a term. For example, we may ask: what determines the meaning of 'cause'?

---

[12] Sider (2011), p. 1.

[13] Sider (2011), pp. 2-3.

[14] See, in particular, Sider (2011), pp. 85 – 104.

To delineate the question that my proposed account of causation aims to answer, it will be useful to begin by considering the metasemantic question of what determines the meaning of 'cause'. One simple and attractive answer to this question is provided by what Sider calls 'simple charity-based descriptivism'.[15] According to simple charity-based descriptivism, there are two factors that play a role in determining the meaning of a term:

The first factor is *charity to use*: when we are considering a community of speakers who use a particular term τ, one of the things we need to take into account when we are trying to decide whether a given interpretation of τ is correct is the extent to which this interpretation makes the speakers' beliefs (or reasonable beliefs) come out true.

The second factor is *eligibility*: when we are trying to decide whether a given interpretation of τ is correct, we should not merely consider charity to use – we should also consider how well the suggested interpretation succeeds at 'carving at the joints'. This second component of simple charity-based descriptivism presupposes the idea that the world is structured, so that – at least in some cases – some suggested interpretations are objectively more joint-carving, and hence more eligible, than others.

Clearly, both charity to use and eligibility come in degrees. According to simple charity-based descriptivism, the correct interpretation of a term τ is simply the interpretation that achieves the best balance between the two.[16]

Simple charity-based descriptivism entails that in some cases the correct interpretation may be one that is, in fact, not very charitable: charity to use and eligibility trade off, and an interpretation that carves much better at the joints than its rivals may therefore win out overall, in spite of the fact that it is not as charitable.[17] Sometimes, this feature of simple charity-based descriptivism leads

[15] Sider (2011), p. 31.

[16] Sider (2011), pp. 31-32. Simple charity-based descriptivism is closely based on Lewis (1983), and (1984). Essentially the same view is defended in Weatherson (2003).

[17] Sider (2011), pp. 31-32.

to counterintuitive results. For this and other reasons, it is likely that simple charity-based descriptivism – at least in the version considered here – should be rejected as a general metasemantic theory.[18]

However, simple charity-based descriptivism does yield exactly the right results when applied to theoretical terms, including natural kind terms, since these terms are *intended* to carve at the joints. To illustrate this point, one may for example consider the case of theoretical terms in physics, such as 'electron'. The term 'electron' is *intended* to carve at the joints. Thus, a highly joint-carving interpretation of 'electron' *should* trump less joint-carving, but more charitable interpretations: when there is a mismatch between the physicists' theory of 'electrons' and the correct theory (which may not have been discovered yet) about whatever is the most joint-carving property in the vicinity, then the intuitively correct thing to say is that the physicists are mistaken – not that they are talking correctly about something else.[19]

It seems reasonable to take 'cause' to be a theoretical term in this sense: it is *intended* to carve at the joints. Indeed, the importance we attribute to the distinction between causes and non-causes – both in philosophical theory and in our everyday lives, where the distinction underlies e.g. attributions of

---

[18] Sider (2011), pp. 32-33.

[19] Sider (2011), p. 32. It may be helpful to think of this in terms of a distinction between first-order and second-order beliefs and intuitions. Consider, for example, knowledge. First-order beliefs and intuitions about knowledge are concerned with the question whether a person knows that *p* in some particular case. Second-order beliefs and intuitions are concerned with the question whether our first-order beliefs and intuitions can be mistaken – for example, whether we can be mistaken in our intuitive judgements about whether a person knows that *p* in some particular case. In the case of physics, we have a strong second-order intuition that physicists *can*, within limits, be mistaken in their first-order beliefs and intuitions about, for example, electrons. Similarly, in the case of knowledge, we have a strong second-order intuition that we *can*, within limits, be mistaken in our first-order judgements about whether a person knows that *p* in a particular case. Accepting simple charity-based descriptivism allows us to respect these second-order intuitions, since simple charity-based descriptivism allows for the possibility that we can, within limits, be mistaken in our use of terms such as 'knows' and 'electron': according to simple charity-based descriptivism, such mistakes occur when there is a highly eligible candidate meaning in the vicinity, and the superior eligibility of this candidate meaning trumps the fact that it makes some of our claims come out false. See Weatherson (2003), pp. 25-26.

responsibility – would be hard to justify if we were not intending our talk about causation to carve at the joints. And if 'cause' is indeed a theoretical term in this sense, then it is reasonable to hold that the meaning of 'cause' is determined according to simple charity-based descriptivism, or a relevantly similar metasemantic theory.

On this view, the correct interpretation of 'cause' is simply the interpretation that strikes the best balance between eligibility and charity to use. My aim in this dissertation will be to answer just this question:

> *The Question*: which interpretation of 'cause' strikes the best balance between eligibility and charity to use?

If simple charity-based descriptivism – or a relevantly similar metasemantic theory – is true, then the answer to this question will, *ipso facto*, be an answer to the semantic question: what does 'cause' mean? In that case, one may think of my project simply as the project of giving a semantics for 'cause'.

It is, however, important to note that the interest and relevance of *The Question* is independent of whether simple charity-based descriptivism – or a relevantly similar metasemantic theory – is true. To see this, suppose that there is some mismatch between our use of the term 'cause' and the most joint-carving relation in the vicinity. And suppose further that the correct metasemantic theory is such that the intended interpretation of our term 'cause' is not at all sensitive to eligibility. In that case, our term 'cause' may have a highly non-joint-carving meaning. And answering *The Question* will not at all help us to determine what that meaning is.

Instead, we may think of *The Question* as being concerned with the semantics of a different term, 'cause*', introduced via the following stipulation:

> 'Let the meaning of 'cause*' be the meaning that would achieve the best balance between eligibility and charity to use, if the word 'cause' were replaced

> by 'cause*' in all sentences believed (or reasonably believed) by competent speakers of English.'[20]

Why should we take an interest in the semantics of 'cause*'? There are two reasons. Firstly and most importantly, the project of producing a semantics for 'cause*' is interesting considered as a metaphysical project. For when we are concerned with the semantics of 'cause*', we are concerned with the question of what is the most joint-carving relation in the vicinity of our ordinary term 'cause'. And answering this question gives us a deeper insight into the structure of reality – into how to carve reality at its joints – which is, arguably, one of the main goals of metaphysical inquiry.[21]

Indeed, considered as a project in metaphysics, the project of producing a semantics for our ordinary term 'cause' has little or no value, *unless* it is presumed that the meaning of our ordinary term 'cause' is somehow responsive to facts about eligibility. For if the meaning of 'cause' is not responsive to such facts, producing a semantics for 'cause' amounts to nothing more than 'producing a semantics for a fragment of English'.[22] And as Hall asks, '[w]hy should scientists, philosophers of science, or metaphysicians care about that?'[23]

Secondly and more speculatively, we may have a more practical reason to take an interest in the semantics of 'cause*'. For even if the meaning of our ordinary term 'cause' in fact differs somewhat from the meaning of 'cause*', the meaning of 'cause*' is arguably what we *should* mean by 'cause', given that the relation expressed by 'cause*' is more joint-carving than the relation expressed by our ordinary term 'cause'. Sider, for example, makes the normative claim that 'it's *better* to think and speak in joint-carving terms'.[24]

---

[20] This stipulation is inspired by a similar stipulation proposed by Sider (2011), p. 76.

[21] Sider (2011), p. 1.

[22] Hall (2007*b*), p. 3.

[23] Hall (2007*b*), p. 3.

[24] Sider (2011), p. 61.

These considerations show that the importance and relevance of *The Question* is independent of whether an answer to *The Question* produces a semantics for our ordinary term 'cause' or only for 'cause*': the project of answering *The Question* is important as a metaphysical project – and, perhaps, as a project in prescriptive semantics.[25]

Since I believe that the meaning of our ordinary English term 'cause' *is* the meaning that strikes the best balance between eligibility and charity to use, so that 'cause' and 'cause*' have precisely the same meaning, I will simply use 'cause' in the following. However, the reader should bear in mind at all times the assumption I have made about the meaning of 'cause', and feel free to replace 'cause' with 'cause*' as appropriate.

2. The shape of the answer: a biconditional

In this dissertation, I shall given an answer to *The Question* in the form of a biconditional. In this section, I set out in more detail what I aim for in providing this answer. I begin (section 2.1) by considering the left-hand side of the biconditional. I then clarify that I aim to provide an ontological reduction of causation (section 2.2), and finally, I set out the degree of necessity with which the biconditional should hold (section 2.3).

2.1 The left-hand side of the biconditional

Before we can write out the left-hand side of the biconditional, a question arises concerning the *adicity* of the causal relation, i.e. the *number* of the causal relata. It is traditionally assumed that causation is a binary relation, and this assumption in turn dictates the traditional formulation of the left-hand side of the biconditional. As Schaffer writes:

---

[25] Although I have relied on semantic and metasemantic considerations to arrive at *The Question*, it is thus a question about what causation is out in the world, in the spirit of Williamson (2007).

> 'Causation is widely assumed to be a binary relation: *c* causes *e*. Anyone familiar with the literature will recognize the pattern. One starts with '*c* causes *e* if and only if . . .' and then considers how to continue.'[26]

However, this assumption of binarity prejudges the issue: prior to looking at the details of which relation strikes the best balance between eligibility and charity to our use of 'cause', there is no compelling reason to assume that causation is a binary relation.

It is indeed true that the surface form of our causal claims suggests that causation is a binary relation: typically, causal claims take the form '*c* causes *e*'. However, one cannot always read off the adicity of a relation from the surface form of our claims about it. To take just one example, consider the relation of simultaneity. The surface form of our claims about simultaneity suggests that simultaneity is a binary relation, since claims about simultaneity take the form '*a* is simultaneous with *b*'. However, the theory of relativity tells us that the relation of simultaneity is in fact ternary – with the third relatum being a frame of reference.[27]

At present, I therefore believe that we should leave all options open in our formulation of the left-hand side of biconditional. For this reason, I take it as my aim to complete the following biconditional, where it is left open whether the causal relation includes further relata:

> *c* is a cause of *e* relative to [. . .] if and only if [. . .]

More precisely, this should be understood as a biconditional schema,[28] where '*c*' and '*e*' are variables standing for the primary causal relata, which we may, for now, think of as token events (for more on the primary causal relata, see Chapter 4). To complete the left-hand side of this biconditional, we then need

---

[26] Schaffer (2005), p. 297.

[27] Cf. Schaffer (2016), p. 17.

[28] Paul and Hall (2013), p. 25.

to determine whether there are any further causal relata – and if so, what they are. I shall argue that there *is* a further relatum – namely, what I call a *possibility horizon* (see Chapter 5).

Note that I use the ordinary English term 'cause' in my statement of the left-hand side of the biconditional. My reason for doing so is that I believe, as discussed in section 1, that the meaning of our ordinary term 'cause' is the meaning that strikes the best balance between eligibility and charity to use. To avoid relying on this assumption, however, one may rewrite the biconditional, replacing 'cause' with the term 'cause*' introduced in section 1:

> *c* is a cause* of *e* relative to [. . .] if and only if [. . .]

As mentioned above, the reader should feel free to replace 'cause' with 'cause*' as appropriate.

2.2 Ontological reduction

A good answer to *The Question* completes the biconditional in a way that is *informative*. Thus, we need to rule out ways of completing the biconditional that are true – even necessarily true – but utterly uninformative, such as the following:

> *c* is a cause of *e* if and only if *e* is an effect of *c*.

In the following, I will set myself a more specific goal, which ensures that the completed biconditional is suitably informative. This goal is based on the assumption that causation is not an irreducible, perfectly fundamental feature of the world. Rather, causation *reduces to* other more fundamental, non-causal features of the world.

While this reductionist assumption is not uncontroversial,[29] it is widely shared in the literature. For example, Hall writes:

> 'a sensible metaphysical position is that facts about what causes what *reduce to* facts about the complete history of physical states the world occupies, together with facts about the fundamental laws that govern the evolution of these states.'[30]

I will not argue for this assumption here, except by attempting to show that the project it prompts us to undertake can be successfully carried out. Plausibly, however, the success of this project is all that is required to vindicate the reductionist assumption.[31]

What is this project? Given the reductionist assumption, a natural aim in completing the biconditional is to show *how* causation reduces to more fundamental, non-causal features of reality. Borrowing a term from Paul and Hall, we may describe an account of causation that is intended to achieve this aim as an *ontological reduction* of causation.[32] The name is apt because it shows where an account with the above aim must be situated relative to two crucial distinctions: the distinction between *ontological* and *conceptual* accounts, and the distinction between *reductive* and *non-reductive* accounts.

It is relatively simple to draw the distinction between *ontological* and *conceptual* accounts of causation: ontological accounts aim to lay bare what causation is out in the world, whereas conceptual accounts (often referred to as *analyses*) aim to illuminate our *concept* of causation.

The distinction between *reductive* and *non-reductive* accounts is harder to pin down. Simply put, an account of causation is reductive just in case it shows

[29] For defences of the anti-reductionist position, see e.g. Armstrong (2004), and Tooley (1990). For a short discussion, see Paul and Hall (2013), pp. 67-69.

[30] Hall (2007*b*), p. 2. Cf. Paul and Hall (2013), pp. 7-8.

[31] Paul and Hall (2013), p. 69.

[32] Paul and Hall (2013), p. 29.

how facts about what causes what obtain in virtue of facts about non-causal features of the world. At the linguistic level, this means that an account of causation is reductive when it provides necessary and jointly sufficient conditions for causation *stated in purely non-causal terms*, where a term is non-causal just in case it refers to a non-causal feature of the world.

To give a full characterisation of the distinction between reductive and non-reductive accounts, however, more needs to be said. For how exactly should we draw the distinction between causal and non-causal terms or, correspondingly, between causal and non-causal features of the world? This question is especially difficult because any feature of the world to which causation is reduced must in *some* sense be causal, in virtue of the very fact that causation reduces to it.[33] Because of these difficulties, I will not attempt to *define* what it takes for a term, or a feature of the world, to be non-causal. Rather, I will rest content with the following incomplete characterisation of what is required from a reductive account:

An account is clearly *not* reductive if it makes use of explicitly causal terms, such as 'cause', 'effect', 'consequence', 'make happen', 'intervene', 'manipulate', etc. This suggests a requirement that a reductive account must satisfy: it must not use terms that feature on the above list.[34]

Sometimes it is easy to see that a proposed account fails to meet this requirement: if one or more of the prohibited terms features on the right-hand side of the biconditional, the proposed account clearly fails to be reductive. In other cases, it may not be immediately obvious whether a proposed account meets the requirement. Consider, for example, an account that reduces causation to, among other things, relations of counterfactual dependence. The term 'depends counterfactually on' is not on the list of prohibited explicitly causal terms. So, at a first glance, it might seem that the proposed account

[33] I am grateful to Katherine Hawley for making me aware of this point.

[34] This requirement is parallel, for example, to van Inwagen's requirement that an answer to his Special Composition Question must not make use of explicitly mereological terms such as 'part', 'whole', 'fusion', etc. See van Inwagen (1990), p. 31.

meets the requirement. To see whether it in fact does so, however, we must ask whether the proposed account can be *completed* without relying on terms from the prohibited list. In this case, the proposed account must, to be complete, include an account of the truth conditions of the counterfactuals it relies on. And if these truth-conditions in turn involve terms from the prohibited list, then the account fails to be reductive.[35]

In many cases, a proposed account of causation can be shown to be non-reductive because it fails to meet the requirement presented here. However, the requirement is open-ended in at least two ways: I have not given a full list of prohibited terms, and I have not said at what point an account can be considered complete. Thus, I have not yet given a characterisation that allows me to show that a proposed account *is* reductive. Paul and Hall offer the following ideal requirement for a reductive account:

> 'show, explicitly, how facts about causation are grounded in facts about fundamental physical states, together with facts about the fundamental physical laws governing their evolution.'[36]

This offers the positive suggestion that facts about fundamental physical states and facts about the fundamental physical laws *should* be allowed into the reductive basis. On this suggestion, then, an account of causation that is based *solely* on these facts counts as being sufficiently reductive.

My account of causation comes close to satisfying this ideal requirement: it is based on four basic ingredients – complete states, laws of nature understood as rules specifying the temporal evolution of complete states, relations of overall similarity between complete states, and the space of metaphysically possible worlds.

[35] Cf. Paul and Hall (2013), p. 38.

[36] Paul and Hall (2013), p. 40.

Finally, it is worth noting that the aim of giving an ontological reduction of causation has implications for the *relation* between the two sides of the biconditional: the aim of giving an ontological reduction requires that the facts about what causes what, stated on the left-hand side of the biconditional, should *reduce to* the non-causal facts stated on the right-hand side.

More generally, I aim to complete the biconditional in such a way that the two sides are not on a par: firstly, I aim to complete the biconditional in such a way that the facts on the right-hand side are ontologically *more fundamental* than the facts on the left-hand side. Secondly and relatedly, I aim to complete the biconditional in such a way that the right-hand side is *explanatorily prior* to the left-hand side: the facts about what causes what, stated on the left-hand side, should be explained by the more fundamental, non-causal facts stated on the right-hand side, and not *vice versa*.

If you like, you may think of the relation between the two sides of the biconditional in terms of grounding, where 'grounding' is used simply as 'a placeholder for some relation of non-causal, ontological dependence'.[37] Given this very minimal notion of grounding, my aim is to complete the biconditional in such a way that the facts about what causes what, stated on the left-hand side, are *grounded in* – *depend on*, or *obtain in virtue of* – the facts stated on the right-hand side.

2.3 Truth in all deterministic worlds

I aim to arrive at a completed biconditional that is true in all worlds governed by forwards deterministic laws that are time-translation invariant, and do not permit action at a temporal distance.[38] We may think of this as an implicit restriction on the biconditional I propose: it applies to worlds whose laws are forwards deterministic, time-translation invariant, and do not permit action at a temporal distance; otherwise, it simply falls silent.

[37] Sartorio (2016), p. 8.

[38] Paul and Hall (2013) adopt similar restrictions (see Paul and Hall (2013), p. 8).

An account of causation that applies only to this class of worlds is obviously incomplete: even though deterministic laws usually give an excellent account of the macroscopic goings-on in the actual world, quantum physics suggests that the actual world is governed by indeterministic laws. To arrive at an account of causation that applies to the actual world, we therefore need an account of causation under indeterministic laws. Given this, why focus on developing an account of causation that is restricted to deterministic worlds?

My reason is that it seems plausible that the very same causal relation is found in deterministic and indeterministic worlds. Given this, it makes sense to begin with the comparatively easier task of arriving at an account of causation that applies only to deterministic worlds: this already imposes sufficient constraints on the proposed account, and it avoids the added complexities that arise when dealing with indeterministic laws.[39] My hope is that my proposed account may then eventually be generalised, so as to apply smoothly to worlds governed by indeterministic as well as deterministic laws.

In addition, it is worth noting that even though I only consider worlds with forwards deterministic laws that are time-translation invariant and do not permit action at a temporal distance, the class of worlds under consideration still includes many very different kinds of worlds. Consider, for example, a world governed by the rules of Conway's Game of Life. Paul and Hall describe such a world as follows:

> 'space is divided up into discrete cells, each of which can either be occupied or unoccupied; time is divided up into discrete moments; the pattern of occupation of the cells at one moment is lawfully and deterministically fixed by the pattern of occupation of the cells at the prior moment.'[40]

---

[39] Cf. Paul and Hall (2013), p. 63.

[40] Paul and Hall (2013), p. 56; cf. Maudlin (2004). For the original description, see Gardner (1970).

Although such a world is very different from our own, we are able to recognise causal relations. And since this world has forwards deterministic laws that are time-translation invariant and do not permit action at a temporal distance, it is included among the worlds under consideration. Thus, my proposed biconditional needs to yield correct verdicts about what causes what in such a world. This already rules out a significant class of accounts of causation, namely accounts that attempt to reduce causation to transfers of conserved quantities:[41] in Conway's Game of Life, there are no transfers of conserved quantities, but there are still causal relations that need to be accommodated by my proposed account.[42]

Finally, my proposed account is only intended to apply to cases where the candidate cause *c* occurs strictly earlier than the effect *e* (within some appropriately chosen frame of reference). In other cases – that is, in putative cases of simultaneous or backwards causation – my account simply does not apply. I remain neutral on the question of whether there *are* genuine cases of simultaneous or backwards causation.[43] Either way, however, I do not need to exclude worlds that permit simultaneous or backwards causation: my account can still be applied in such worlds – though only to those cases where the candidate cause *c* occurs strictly earlier than the effect *e*.[44] If there are cases of simultaneous or backwards causation, I leave aside the question of whether my account of causation can be generalised to cover such cases, or they require a separate treatment.

[41] Such as the accounts of Dowe (2000); Fair (1979); and Salmon (1994).

[42] Cf. Paul and Hall (2013), p. 56.

[43] For a brief discussion, see Collins, Hall, and Paul (2004*b*), pp. 9-12.

[44] By contrast, Paul and Hall impose the further restriction on the class of worlds under consideration that the laws must not permit backwards causation (see Paul and Hall (2013), p. 8. I have chosen the different option outlined above for two reasons: first, merely limiting the questions to which my account applies, rather than limiting the class of worlds, ensures that my account applies more widely. Second, it is not clear that a prohibition against laws permitting simultaneous or backwards causation can be stated without appealing to implicitly or explicitly causal terms.

3. Conclusion

To sum up, we may give the following characterisation of the aim of my proposed account of causation:

> My aim is to provide an account of causation, presented as a biconditional,
>
> *c* is a cause of *e* relative to […] if and only if […]
>
> where the meaning of the word 'cause' on the left-hand side is the meaning that strikes the best balance between eligibility and charity to use, and where the biconditional is completed in such a way that
>
> 1) it provides an ontological reduction of causation, showing how causation reduces to more fundamental, non-causal features of the world, and
> 2) it is true in all instances where *c* occurs strictly earlier than *e* (within some appropriately chosen frame of reference), and where the relevant world is governed by forwards deterministic laws that are time-translation invariant and do not permit action at a temporal distance.

2

The dual nature of causation

In this chapter I give an overview of my proposed account of causation. The key to this account is the idea that causation has a dual nature: on the one hand, a cause must be connected to its effect via a *genuine process*; on the other hand, a cause must *make a difference* to its effect.

As I have set out in detail in Chapter 1, my aim is to identify the best candidate meaning of 'cause' – that is, the relation out in the world that strikes the best balance between eligibility and charity to our use of 'cause'. To identify this relation, we need to pay close attention to our use of 'cause' as it is revealed in the pre-theoretic causal judgements of competent speakers. In the following, I will use 'intuition', 'intuitive', 'intuitively', etc., to mark such judgements.

Given that my aim is to find the relation out in the world that strikes the best balance between eligibility and charity to our use of 'cause', intuitions should not be treated as non-negotiable 'data':[1] we need to leave open the possibility that the best candidate meaning of 'cause' is one whose imperfect charity to use is outweighed by its superior eligibility. Rather, we may think of our intuitions as well-intentioned, but perhaps incomplete, reports of the causal relation out in the world – and our task now is to use these reports as guides to find out what the causal relation out in the world really is.

In the following, I will argue that a close study of our intuitive judgements – and, in particular, of the *tensions* between different intuitive judgements – suggests that causation has a dual nature. To bring this out, I

[1] Cf. Hall (2007*b*), p. 2.

begin (in section 1) by giving a brief characterisation of the intuitive judgements that characterise our use of 'cause' and, in particular, of the tensions between them. Next, I argue (in section 2) that these tensions suggest that the causal relation itself has a dual nature.

1. Tales of the beast: our use of 'cause'

Intuitions about causation may be divided into two kinds: intuitions about specific cases, and intuitions about general principles.

By now, the literature on causation contains a multitude of specific cases where we have firm intuitions and widespread agreement about what causes what. These intuitive judgements about specific cases are our main guides to the causal relation out in the world. And the degree to which a proposed candidate meaning of 'cause' can accommodate these intuitive judgements is the main measure of its charity to use.

The second kind of intuitions about causation deal with general principles, such as the intuitive principle that counterfactual dependence is sufficient for causation, or that causation is a transitive relation. Such principles sanction particular causal inferences. For example, the principle that counterfactual dependence is sufficient for causation sanctions the inference from '*e* depends counterfactually on *c*' to '*c* is a cause of *e*', and the principle that causation is a transitive relation sanctions the inference from '*c* is a cause of *d*' and '*d* is a cause of *e*' to '*c* is a cause of *e*'. These principles summarise key features of our practices of causal inference, and I believe that we should consider these practices to be an integral part of our use of 'cause'.

To give a clear presentation of specific cases, I will frequently rely on neuron diagrams. In section 1.1, I set out the conventions governing the interpretation of neuron diagrams, leaving an overview over all cases discussed in this dissertation for Appendix B.

Concerning intuitive principles of causal inference, I shall focus on what I take to be the three most important principles: the intuitive principle that counterfactual dependence is sufficient for causation, the intuitive principle that causation is a transitive relation, and the intuitive principle that causation is intrinsic to a process. I present these intuitive principles in section 1.2. Finally, I show (in section 1.3) that there are deep tensions between the three intuitive principles and our intuitions about specific cases.

1.1 Specific cases: representation using neuron diagrams

Neuron diagrams do an excellent job of 'representing a complex situation clearly and forcefully, allowing the reader to take in at a glance its central causal characteristics'.[2] For this reason, I will frequently use neuron diagrams to represent specific cases. In this section, I introduce the conventions governing the interpretation of neuron diagrams, and illustrate how we may use neuron diagrams to represent the causal structure of real life situations.[3]

Let us begin by considering the following neuron diagram, illustrating the structure of a standard case of early preemption:

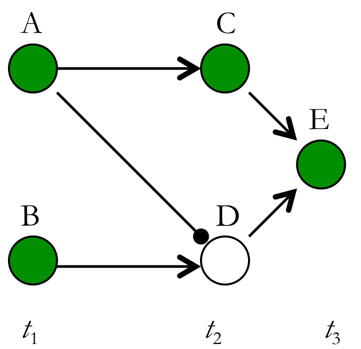


Figure 1

Figure 1 shows a system of neurons, including the stimulatory and inhibitory channels between them.[4] Neurons are indicated by circles. If a neuron fires, it is coloured green; if it does not fire, it is left blank. Stimulatory channels between neurons are indicated with arrows; inhibitory channels are indicated with a line ending with a blob.

[2] Paul and Hall (2013), p. 10.

[3] I discuss the interpretation of neuron diagrams further in Chapter 3 section 4, and add conventions for naming neuron events in Chapter 4 section 2.3.

[4] In the following, I number figures consecutively within each chapter.

The neuron diagram should be read from left to right. On the far left is shown what happens at time t_1: neurons **A** and **B** both fire. Further to the right is shown what happens at the later time t_2: neuron **C** fires, while neuron **D** fails to fire. And furthest to the right is shown what happens at the still later time t_3: neuron **E** fires. In the following, I use italicized letters to denote the firings of neurons. Thus, '*A*' denotes the firing of **A**, '*B*' denotes the firing of **B**, etc.

It is part of the conventions for interpreting neuron diagrams that we assume, unless an explicit stipulation is made to the contrary, that the pattern of firings evolves forward in accordance with the following neuron laws:

i. The firing of a neuron lasts only an instant.
ii. When a neuron fires, it sends signals through its out-going stimulatory and inhibitory channels; otherwise, it sends no signals.
iii. A signal sent at time t_i arrives at the next neuron at time t_{i+1}.
iv. For any time t, if a neuron receives no signals or receives at least one inhibitory signal at time t, it does not fire at t.
v. For any time t, if a neuron receives at least one stimulatory signal and no inhibitory signals at t, it fires at t.

To illustrate, we may now see how these conventions play out in the neuron diagram above: **A** fires at time t_1, and sends signals through its out-going stimulatory and inhibitory channels. Its stimulatory signal reaches **C** at time t_2, and its inhibitory signal reaches **D** at time t_2. In addition, **B** fires at time t_1 and sends a stimulatory signal, which reaches **D** at time t_2. At time t_2, **C** thus receives a stimulatory signal (from **A**) and no inhibitory signals, and **C** therefore fires at time t_2. By contrast, **D** receives an inhibitory signal (from **A**) at time t_2, and **D** therefore does not fire at time t_2. Finally, the stimulatory signal from **C** reaches **E** at time t_3, and since **E** receives no inhibitory signals at that time, **E** therefore fires at time t_3.

The conventions stated above cover most of the cases I will discuss in this dissertation. In a few cases, however, I use neuron diagrams with special features. For example, I sometimes consider stubborn neurons that require two stimulatory signals in order to fire, or neurons that can fire in two or more different ways – for example, either in uniform green or in stripes. When I do so, I explicitly state how the conventions above should be extended to cover the case at hand.

Finally, I adopt the convention of treating the information contained in a neuron diagram as *complete*, in the sense that there is no hidden information that may overturn the judgements we reach by considering what is explicitly represented in the neuron diagram.

As I will now illustrate, neuron diagrams may be used to give a perspicuous representation of real life cases. For example, Figure 1 above may be used to represent the structure of the following real-life situation:

> *Early preemption*: Suzy throws a rock at a window and the window shatters. If Suzy had not thrown, Billy – who is standing right next to her – would have thrown his rock, and the window would still have shattered.

Figure 1 represents the structure of *Early preemption* as follows: *A* (i.e. the firing of neuron **A**) represents Suzy's throw, *C* represents Suzy's rock flying towards the window, and *E* represents the shattering of the window. Furthermore, *B* represents Billy's readiness to throw, and **D**'s failure to fire represents Billy's failure to throw.

Furthermore, the pattern of stimulatory and inhibitory channels between the neurons represents (simplified versions of) the nomological relationships between these events: Suzy's rock flies towards the window at t_2 if and only if Suzy throws at t_1; Billy throws his rock at t_2 if and only if he is ready to throw at t_1 *and* Suzy does not throw at t_1; and the window shatters at t_3 if and only if Suzy's rock is flying towards it at t_2 *or* Billy throws at t_2.

The pattern of stimulatory and inhibitory channels also captures the counterfactual structure of the case: in the real life situation, the window would still have shattered if Suzy had not thrown – because of Billy's throw. Correspondingly, the neuron laws ensure that *E* would still have occurred even if *A* had not occurred, as illustrated below:

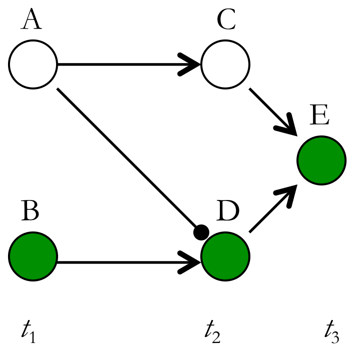


Figure 1*

Throughout this dissertation, we shall look at a variety of real-life cases and use neuron diagrams to bring out their structure. In Appendix B, I give an overview of all the neuron diagrams considered in this dissertation.

1.2 Intuitive principles of causal inference: sufficiency of counterfactual dependence, transitivity, and intrinsicness

In the following, I will focus on three general principles that capture our most central practices of causal inference.[5] These three principles may be summed up as follows:

I. Counterfactual dependence is sufficient for causation.
II. Causation is a transitive relation.
III. Causation is intrinsic to a causal process.

[5] These three principles are among the basic claims about causation listed in Paul and Hall (2013), p. 4.

Each of these principles sanctions a particular kind of causal inference. In this section, I will say more about how we should understand each of these principles and, correspondingly, what the associated causal inference is.

Let us begin with the first principle – that counterfactual dependence is sufficient for causation. In Chapter 8, I discuss the evaluation of the relevant counterfactuals in detail. For now, however, we may rely on our intuitive understanding of counterfactuals. Based on this, we may give the following preliminary statement of the principle, making it more perspicuous what kind of inference it sanctions (here and in my statement of the following two principles, the asterisk indicates that I do not endorse this principle):

> **Sufficiency of counterfactual dependence for causation*:
> if *e* depends counterfactually on *c*, then *c* causes *e*.

The second principle states that causation is a transitive relation. We may restate this principle as follows, to clarify the kind of inference it sanctions:

> **Transitivity of causation*: if there is a set of events $\{d_1, d_2, \ldots, d_n\}$, such that *c* causes d_1, d_1 causes d_2, . . . , and d_n causes *e*, then *c* causes *e*.[6]

Finally, the third principle – that causation is intrinsic to a causal process – requires a little more explanation. It should be understood roughly as follows: when *c* causes *e*, this relation holds in virtue of the laws of nature together with the intrinsic character of *c*, *e*, and the process connecting *c* to *e* – any other features of the environment are irrelevant. Lewis spells out and motivates this principle as follows:

[6] This principle is equivalent to the following simpler principle: if *c* causes *d* and *d* causes *e*, then *c* causes *e*. I have chosen the above statement of the principle because this makes things go more smoothly when I present a restricted principle of transitivity in Chapter 11.

> 'Suppose we have processes – courses of events, which may or may not be causally connected – going on in two distinct spatiotemporal regions, regions of the same or of different possible worlds. Disregarding the surroundings of the two regions, and disregarding any irrelevant events that may be occurring in either region without being part of the process in question, what goes on in the two regions is exactly alike. Suppose further that the laws of nature that govern the two regions are exactly the same. Then can it be that we have a causal process in one of the regions but not in the other? It seems not. Intuitively, whether the process going on in a region is causal depends only on the intrinsic character of the process itself, and on the relevant laws. The surroundings, and even other events in the region, are irrelevant.'[7]

Lewis's statement of this principle is objectionably vague. In particular, he does not say what it takes for an event to be part of a process.[8] For now, however, we may simply rely on our intuitive understanding of what it takes for an event to be part of a process. Based on this, we may give a preliminary statement of the principle as follows:

> **Intrinsicness of causation*: if c causes e, and a structure of events $\mathscr{S}$, including all the events that are involved in a process connecting c to e, is governed by the same laws and exactly matches a structure of events $\mathscr{S}^*$, then the counterpart c^* of c in $\mathscr{S}^*$ is a cause of the counterpart e^* of e in $\mathscr{S}^*$.

In Chapter 11 section 3, I show how we can make this precise.[9]

---

[7] Lewis (1986*e*), p. 205.

[8] Hall (2004*b*), p. 239; Paul and Hall (2013), p. 126.

[9] For further discussion of the intrinsicness principle, see Hall (2002) and (2004*b*); Paul and Hall (2013), pp. 124-131; and Sartorio (2016), pp. 71-75.

1.3 Tensions between the intuitive principles and our intuitions about specific cases

Our intuitions about specific cases come into *prima facie* conflict with each of the three general principles set out above. In the following, I will consider each principle in turn and show how such conflicts arise.

1.3.1 Counterexamples to the sufficiency of counterfactual dependence for causation

In this section, I consider two kinds of cases that both present apparent counterexamples to the sufficiency of counterfactual dependence for causation: cases of omission-involving causation, and cases where the candidate cause is, so to speak, at the wrong level of detail relative to the effect. I suggest that cases of omission-involving causation do not present genuine counterexamples to the sufficiency of counterfactual dependence, while cases involving mismatch in the level of detail do.

Cases involving omissions are often cited as counterexamples to the sufficiency of counterfactual dependence for causation.[10] Consider, for example, the following case:

> *The flowers*: Suzy goes on holiday and Billy promises to water her flowers while she is away. However, Billy does not water the flowers and the flowers die.

In this case, the death of the flowers depends counterfactually on Billy's failure to water them: if Billy had not failed to water the flowers, they would not have died. Correspondingly, we intuitively judge that Billy's failure to water the flowers is a cause of their death. So far, there is no counterexample to the sufficiency of counterfactual dependence for causation.

However, the death of the flowers also depends counterfactually on other people's failures to water them – for example, the death of the flowers

[10] See e.g. Beebee (2004), pp. 294-95; McGrath (2005), pp. 126-30; Sartorio (2010), pp. 262-3; Thomson (2003), pp. 95-98.

depends counterfactually on the queen of England's[11] failure to water them: if she had watered the flowers, they would not have died. Intuitively, however, the queen's failure to water the flowers is *not* a cause of their death.[12] This is sometimes called the *problem of profligate omissions*: if we accept omissions as causal relata, and we hold that counterfactual dependence is sufficient for causation, then we end up with many more omissions as causes than we would normally accept.[13]

I believe that we should accept the intuitive verdict that the queen's failure to water the flowers is not a cause of their death. Even so, I will argue that cases such as this do not present counterexamples to the sufficiency of counterfactual dependence for causation: as I suggest in Chapter 8, what matters for causation is a special kind of counterfactual dependence – namely, counterfactual dependence *within a possibility horizon*. This notion of counterfactual dependence is inextricably bound up with a notion of *relevant* possibilities: a possibility horizon consists in a restricted class of possible worlds containing only those worlds that represent relevant possibilities. Since the queen's watering the flowers is not a relevant possibility – as Schaffer writes, 'we never supposed that the queen would deign to water [the] flowers'[14] – the queen's watering the flowers is treated as *impossible* within the contextually relevant possibility horizon. And so we find that, within the contextually relevant possibility horizon, the death of the flowers does *not* depend counterfactually on the queen's failure to water them (for my treatment of this case, see Chapter 10 section 1).

There are, however, cases of a different kind that I believe do put pressure on the principle that counterfactual dependence (even when this is understood as counterfactual dependence within a possibility horizon) is sufficient for causation. These are cases where the effect depends

[11] More precisely, the queen of the United Kingdom, Canada, Australia and New Zealand.

[12] See e.g. Sartorio (2010), pp. 262-3; Schaffer (2005), p. 300.

[13] See e.g. Bernstein (2014), p. 429.

[14] Schaffer (2005), p. 302.

counterfactually on an event that, intuitively, is at the wrong level of detail. As a first example, consider the following case:

> *Scarlet*: The pigeon Sophia has been conditioned to peck at scarlet to the exclusion of all other colours. She is presented with a scarlet triangle and pecks at it.[15]

Sophia's pecking depends counterfactually on the triangle's being scarlet: if the triangle had not been scarlet, Sophia would not have pecked. Correspondingly, we intuitively judge that the triangle's being scarlet is a cause of Sophia's pecking. So far, there is no counterexample to the sufficiency of counterfactual dependence.

But now consider the event of the triangle's being red. Once again, we have counterfactual dependence: if the triangle had not been red, Sophia would not have pecked – for in any world in which the triangle is non-red, it is also non-scarlet, and Sophia pecks only at scarlet. Should we therefore say that the triangle's being red is a cause of Sophia's pecking? Doing so seems at best strained: intuitively, the *mere* event of the triangle's being red is not a cause – the cause is the triangle's being scarlet.[16] We thus find a *prima facie* tension between our intuitions about this specific case and the intuitive principle that counterfactual dependence is sufficient for causation.

In the above case, the problem is that the effect depends counterfactually on an event that is, intuitively, at a too coarse level of detail to fit the effect. We may similarly construct cases in which the effect depends counterfactually on an event that is, intuitively, at a too fine level of detail to fit the effect. Consider, for example, the following case:

---

[15] This case is closely based on a case presented in Yablo (1992*a*), p. 257. For similar cases, see Yablo (1992*b*), p. 415, and Sartorio (2010), pp. 266-69.

[16] Cf. Yablo (1992*b*), p. 415; Sartorio (2010), pp. 266-69.

> *Red*: The pigeon Delia has been conditioned to peck at red to the exclusion of all other colours. She is presented with a scarlet triangle and pecks at it. In the lab where she is, the researchers use just two colours – scarlet and emerald. If Delia had not been presented with a scarlet triangle, she would have been presented with an emerald triangle.[17]

In this case, Delia's pecking depends counterfactually on the triangle's being red. Correspondingly, we intuitively judge that the triangle's being red is a cause of Delia's pecking. So far, no counterexample. However, Delia's pecking also depends counterfactually on the triangle's being scarlet – for if the triangle had not been scarlet, it would have been emerald, and Delia does not peck at emerald. Should we therefore say that the triangle's being scarlet is a cause of Delia's pecking? Once again, this seems at best strained: intuitively, the precise shade of red does not matter – all that matters is that the triangle is *some* shade of red.[18]

In the following quotation, I believe that Yablo succeeds in capturing the motivation behind these judgements:

> 'causes are expected to be *commensurate* with their effects: roughly, they should incorporate a good deal of causally important material but not too much that is causally unimportant. […] Although determinables and determinates do not compete for causal *influence*, broadly conceived as encompassing everything from causal relevance to causal sufficiency, they *do* compete for the role of *cause*, with the more commensurate candidate prevailing.'[19]

---

[17] This case is closely based on a case presented in Yablo (1992*a*), p. 257. For similar cases, see Yablo (1992*b*), p. 417, and Sartorio (2010), pp. 264-66.

[18] Cf. Yablo (1992*b*), p. 417; Sartorio (2010), pp. 264-66.

[19] Yablo (1992*a*), pp. 273-4; cf. Yablo (1992*b*). In the literature, the relation between cause and effect that Yablo refers to as 'commensuration' is often called 'proportionality'. For discussion, see Bernstein (2014); Dowe (2010); McDonnell (forthcoming); Shapiro and Sober (2012); and Weslake (2013).

In the above cases, Yablo's suggestion is that the determinate, i.e. the triangle's being *scarlet*, and the determinable, i.e. the triangle's being *red*, compete for the role of cause. In *Scarlet*, the triangle's being scarlet is the more commensurate candidate. Thus, we happily say that the triangle's being scarlet causes Sophia's pecking, while we are reluctant to say that the triangle's being red does so. In *Red*, by contrast, the triangle's being red is the more commensurate candidate. And so, we happily say that the triangle's being red causes Delia's pecking, while we are reluctant to say that the triangle's being scarlet does so.

1.3.2 Counterexamples to the transitivity of causation

There is a range of specific cases that present counterexamples to the intuitive principle that causation is transitive. The following is a typical and intuitively convincing case:

> *Boulder*: 'A boulder is dislodged and begins rolling ominously toward Hiker. Before it reaches him, Hiker sees the boulder and ducks. The boulder sails harmlessly over his head with nary a centimetre to spare. Hiker survives his ordeal.'[20]

Intuitively, the boulder's fall causes Hiker's duck, and Hiker's duck in turn causes his survival. If causation is transitive, it follows that the boulder's fall causes Hiker's survival. But intuitively, this is false: the boulder's fall does *not* cause Hiker's survival. Rather, Hiker survives *in spite of* the boulder's fall. This case – and others like it – shows that there is a *prima facie* tension between our intuitive judgements in specific cases and the intuitive principle that causation is transitive.

[20] Paul and Hall (2013), p. 222. The case is originally due to Hall. For further discussion, see e.g. Hall (2004*a*); Hitchcock (2001); Kvart (1991); Lee (1988); McDermott (1995); Paul (2004*a*); Sartorio (2005); and Schaffer (2005).

1.3.3 Counterexamples to the intrinsicness of causation

Finally, there are specific cases that present counterexamples to the intuitive principle that causation is intrinsic to a process. Consider, for example, the neuron diagram below, which shows a standard case of double prevention:

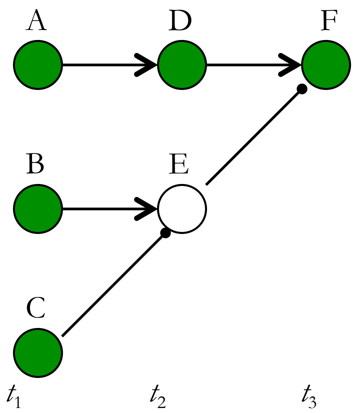


Figure 2

Intuitively, *C* is here a cause of *F*. Furthermore, it is clear that on any reasonable way of drawing the distinction between events that are and are not part of the process leading up to *F*, the firing of **B** certainly is *not* part of that process. On any reasonable way of spelling out the intrinsicness principle, the process from *C* to *F* is therefore exactly the same in Figure 2* below, where neuron **B** fails to fire:

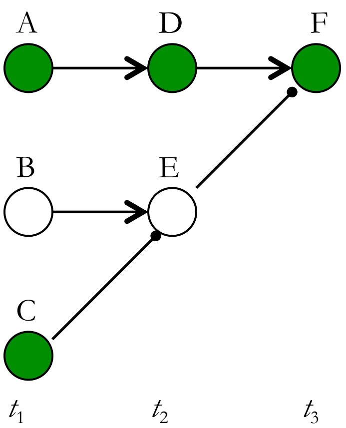


Figure 2*

By the intrinsicness thesis, it now follows that *C* is here a cause of *F*. Intuitively, however, this is simply false: as Paul and Hall note, letting **B** be dormant 'completely reverses the intuitive verdict about *C* – once we remove the threat created by **B**'s firing, we intuitively judge that *C* is completely idle with respect to [*F*]'.[21] Thus, the present case – and others like it – shows that there is a *prima facie* tension between our intuitions about specific cases and the general principle that causation is intrinsic to a process.

1.4 Consequences of the tensions

The above discussion shows that our intuitions about specific cases come into *prima facie* conflict with the three intuitive principles. This shows that it is impossible to find a candidate meaning of 'cause' that is *perfectly* charitable to use: any candidate meaning that perfectly respects our intuitions about specific cases will be imperfectly charitable to use by rejecting the three intuitive principles of causal inference, and *vice versa*. We must, inevitably, compromise.

However, the tensions do not merely lead to this negative conclusion. They also point to an important feature of the casual relation: as I shall argue in the following section, the best explanation of our conflicting intuitions is that the causal relation has a dual nature. This idea allows us to find a candidate meaning of 'cause' that achieves a very high degree of charity to use: the three intuitive principles in their *unrestricted* form come into conflict with our intuitions about specific cases. However, by recognising the dual nature of

[21] Paul and Hall (2013), p. 197. Paul and Hall are here commenting on a version of the case where neuron **B** is not merely dormant, but has been removed entirely. However, their reasoning applies equally well to the case presented in Figure 2*, since letting **B** be dormant is enough to remove the threat created by **B**'s firing. In the following, I adopt a conception of sameness of laws such that two neuron worlds count as being governed by the same laws only if they contain the same structure of neurons, with the same patterns of stimulatory and inhibitory channels (see Chapter 3 section 4). Such a conception is certainly not mandatory, but it makes it much easier to work with neuron diagrams. And given this conception of sameness of laws, the case considered by Paul and Hall does not present a counterexample to the intrinsicness of causation. By contrast, the case considered above, where **B** is simply dormant, does present a counterexample.

causation, we are able to recover *restricted* versions of all three principles, where these restricted principles are fully compatible with our intuitions about specific cases.

2. The beast is a butterfly: the dual nature of causation

The most detailed development of the idea that causation has a dual nature is due to Hall, who has argued that we have two concepts of causation – namely, the concepts of production and dependence. I begin by presenting an outline of Hall's account (section 2.1), as well as the challenges for this account (section 2.2). Building on this, I then present my own picture of the dual nature of causation (section 2.3). And finally, I give a preview of how I will develop my account in subsequent chapters (section 2.4).

2.1 Hall's two concepts of causation: production and dependence

In his paper, 'Two concepts of causation', Hall makes an intriguing suggestion as to how we may resolve the tensions between our different intuitions about causation. His suggestion is that we have *two* concepts of causation: the concept of *production*, and the concept of *dependence*.[22] Dependence is simply counterfactual dependence. Production is, in Hall's words, 'rather more difficult to characterize, but we evoke it when we say of an event *c* that it helps to *generate* or *bring about* or *produce* another event *e*'.[23]

In most of the cases that we intuitively recognise as cases of causation, the two concepts overlap.[24] Sometimes, however, they come apart. And paying close attention to the cases where they come apart is a useful first step towards characterising the two concepts more fully.

Let us begin by considering a case where the cause *produces* the effect, but where the effect does not *depend* on the cause. Indeed, we have already seen an

---

[22] Hall (2004*b*), pp. 252-4.

[23] Hall (2004*b*), p. 225. For Hall's discussion of how to analyse production, see Hall (2004*b*), pp. 257-65.

[24] Hall (2004*b*), p. 254.

example of such a case in section 1.1, namely our standard case of early preemption:

> *Early preemption*: Suzy throws a rock at a window and the window shatters. If Suzy had not thrown, Billy – who is standing right next to her – would have thrown his rock, and the window would still have shattered.

The structure of this case is illustrated in the neuron diagram below:

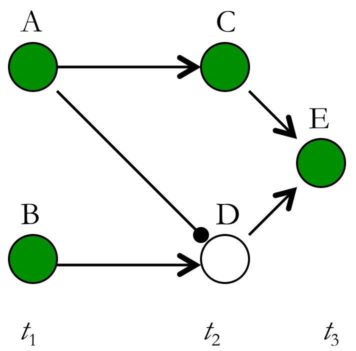


Figure 1

Intuitively, Suzy's throw (*A*) is a cause of the window-shattering (*E*). On Hall's suggestion, Suzy's throw *produces* the shattering – we can trace a productive process from Suzy's throw (*A*), through the flight of her rock (*C*), to the window-shattering (*E*). However, the shattering does not *depend* on Suzy's throw: if Suzy had not thrown, the window would still have shattered – because of Billy's readiness to throw (*B*). Indeed, Hall suggests that production and dependence come apart in this way in all cases of redundant causation.[25]

Next, let us consider a case where the effect *depends* on its cause, but where the cause does not *produce* the effect. Consider, for example, our case of double prevention discussed in section 1.3.3 above:

[25] Hall (2004*b*), p. 253.

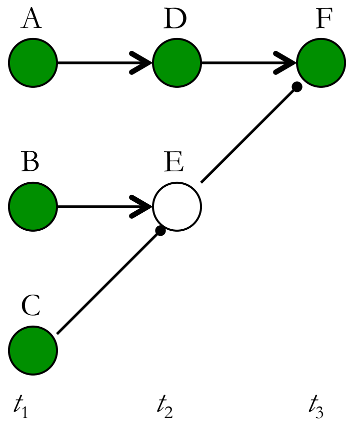


Figure 2

Intuitively, *C* is here a cause of *F*. And in this case, *F depends* on *C*: if *C* had not occurred, *F* would not have occurred. However, *C* does not *produce F* – for, on Hall's conception, production cannot involve omissions such as **E**'s failure to fire.[26]

Considering these cases gives an approximate understanding of the two concepts. We may make this more precise by setting out their general features. Firstly, the two concepts differ – as I hinted at above – in their treatment of omissions: production cannot involve omissions, whereas dependence can.[27] Furthermore, the two concepts differ in their relation to each of the three intuitive principles of causal inference: counterfactual dependence is not sufficient for production, but production is both transitive and intrinsic. By contrast, counterfactual dependence *is*, obviously, sufficient for dependence, but dependence is neither transitive nor intrinsic.[28]

Hall's proposal is that *c* is a cause of *e* just in case *c* produces *e* or *e* depends on *c*. The figure below gives a visual representation of this proposal, where causation is to be identified with the union of production and dependence:

[26] Hall (2004*b*), pp. 253-4.
[27] Hall (2004*b*), p. 226 and 254.
[28] Hall (2004*b*), p. 226 and 253.

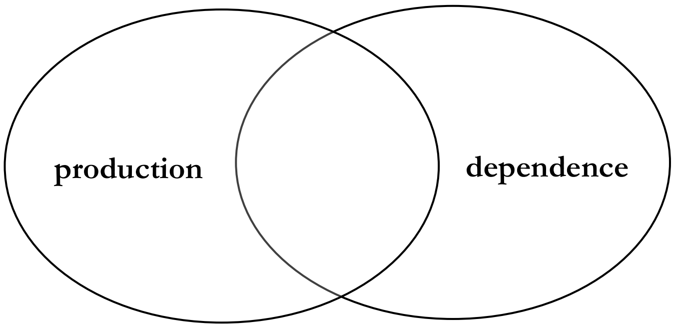


Figure 3

Hall's two concepts proposal is very attractive: it captures a deep truth about our way of thinking about causation as being concerned with, on the one hand, productive processes and, on the other hand, relations of dependence. Furthermore, it shows an intriguing strategy for accommodating our intuitions about both specific cases and general principles: Hall's proposal succeeds in accommodating nearly all of our intuitions about specific cases. Since these intuitions conflict with the three intuitive principles, this means that Hall's proposal cannot accommodate these principles as fully general principles about causation itself. However, it can accommodate each of the three principles as a principle about either production or dependence: counterfactual dependence is sufficient (indeed, both necessary and sufficient) for dependence, but not for production; and production is transitive and intrinsic, whereas dependence is neither.

However, problems arise if we try to use Hall's proposal to determine the best candidate meaning of 'cause'. I turn to these problems in the following section.

2.2 Difficulties for the two concepts account

There is an obvious difference between my project and the project Hall is engaged in: the two concepts proposal targets our *concept(s)* of causation. By contrast, my aim is to capture the relation(s) out in the world that provide the best candidate meaning(s) of 'cause'. These are two very different projects. However, we may still draw on Hall's proposal in trying to determine the best candidate meaning(s) of 'cause'.

Directly translating Hall's proposal yields something like the following: there are two relations out in the world – the relation of *production* and the relation of *dependence*. Dependence is simply counterfactual dependence. Production is harder to characterise. However we may at least say the following: counterfactual dependence is not sufficient for production, production is transitive and intrinsic to a process, and production cannot involve omissions.

Building on this, we now have a choice between the following two options: either we may say that there is a single best candidate meaning of 'cause', namely the disjunction of production and dependence:

> *c* is a cause of *e* if and only if *c* produces *e* or *e* depends on *c*.

Or, alternatively, we may say that our word 'cause' turns out, surprisingly, to be ambiguous between two different relations – what we might call 'production-causation' and 'dependence-causation', where

> *c* is a production-cause of *e* if and only if *c* produces *e*, and
> *c* is a dependence-cause of *e* if and only if *e* depends on *c*.

Given the aim of finding the best candidate meaning(s) of 'cause', the choice between these two options presents a dilemma: the first option does well in terms of respecting the higher-level intuition that there is a *single* meaning of 'cause'. If we take this option, however, we find that 'cause' refers to a disjunctive and therefore not very eligible relation. The second option, by contrast, goes against the higher-level intuition that there is a single meaning of 'cause' and delivers the surprising result that 'cause' is ambiguous between two quite different meanings. However, it then has the advantage that the two relations of production-causation and dependence-causation are both more eligible than the disjunctive relation of production-or-dependence.

Whichever way we resolve this dilemma, there is a price to pay. In addition, the two concepts account faces a further difficulty: as Hall himself recognises, the relations of production and dependence do not capture our intuitions about certain specific cases – in some cases, the proposal is too permissive, i.e. it admits causes that we do not intuitive recognise as such; in other cases, it is not permissive enough, i.e. it rejects causes that we intuitively do recognise.

2.2.1 Counterexamples to the sufficiency of the two concepts account

Let me begin by considering cases in which the two concepts account is too permissive. These cases may be divided into two groups:

i) cases where an event *e depends* on an earlier event *c*,
but where we are reluctant to accept *c* as a cause of *e*, and

ii) cases where *c produces e*,
but where we are reluctant to accept *c* as a cause of *e*.

In section 1.3.1 above, we have already encountered cases belonging to the first group – namely, *Scarlet* and *Red*, which provide counterexamples to the sufficiency of counterfactual dependence for causation. In the following, I will now consider a case belonging to the second group, where *c* produces *e*, but where we intuitively judge that *c* does not cause *e*:

> *Switch*: Suzy is standing by a switch in the railroad tracks. She sees a train approaching in the distance, and flips the switch so that the train travels down the left-hand track. If she had not flipped the switch, the train would have

> travelled down the right-hand track instead. Since the tracks converge a few miles later, the train arrives at its destination all the same.[29]

The structure of the case is represented in the neuron diagram below:

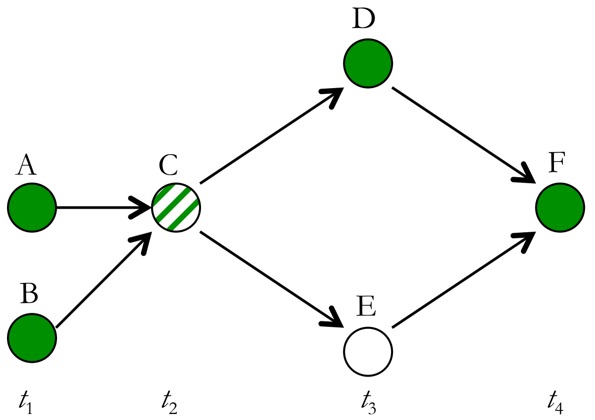


Figure 4

A here corresponds to the train's approach, *B* corresponds to Suzy's flipping the switch, and **C**'s firing corresponds to the train's reaching the place where the track forks. Importantly, **C** can fire in two different ways: in stripes and in uniform green. These two ways correspond to the different settings of the switch: **C**'s firing in stripes represents the train being directed towards the left-hand track; **C**'s firing in uniform green represents the train being directed towards the right-hand track. Finally, *D* represents the train's journey along the left-hand track, **E**'s failure to fire represents the train's failure to journey along the right-hand track, and *F* represents the train's arrival at its destination.

The conventions governing the neuron diagram are exactly as set out in section 1.1 above, with the following important addition to the neuron laws: **C** fires if and only if it receives a stimulatory signal from **A**. Furthermore, the *way* in which **C** fires is determined by whether or not **B** fires: **C** fires in stripes if

[29] Switching cases of this kind are discussed e.g. in Hall (2004*a*), pp. 187-92, (2007*a*), p. 118, and (2007*b*), pp. 28-32 and 51; Hitchcock (2009), p. 394; Mackie (1992); p. 496-7; Paul and Hall (2013), p. 232; and Sartorio (2005), pp. 74-5.

and only if **A** and **B** both fire, and **C** fires in uniform green if and only if **A** fires and **B** does not.

Intuitively, *B* (Suzy's flipping the switch) is *not* a cause of *F* (the train's arrival). Rather, *B* is simply *irrelevant* to *F*. However, from the transitivity and intrinsicness of production it follows that *B produces F*. As Hall himself notes, his account thus has the undesirable consequence that we 'must call switches *producers* of the relevant effect, and so in one central sense *causes*'.[30]

2.2.2 Counterexample to the necessity of the two concepts account

As Hall himself shows, the two concepts account is put under further pressure by cases indicating that it is not permissive enough. Consider, for example, the case below: [31]

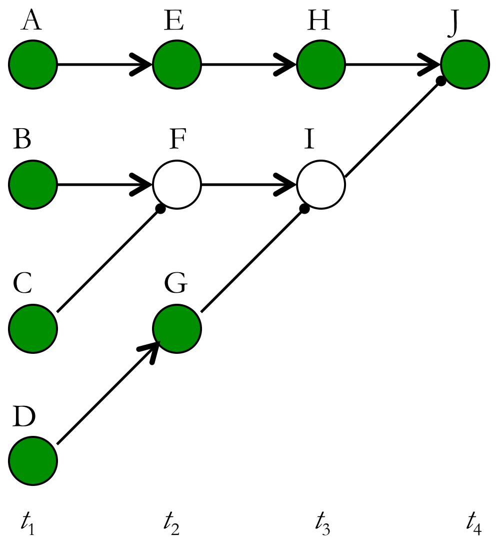


Figure 5

Intuitively, there is a clear sense in which *C* is a cause of *J*: as Hall writes, '[*D*] notwithstanding, it is *C* that *in fact* cancels the threat to [*J*], and cancelling a threat is one way to be a cause'.[32] On the two concepts account, however, there is *no* sense in which *C* is a cause of *J*: *C* does not *produce J*, since production on

[30] Hall (2007*b*), p. 51.

[31] Figure from Hall (2007*b*), p. 52.

[32] Hall (2007*b*), p. 52.

Hall's conception cannot involve omissions. And *J* does not *depend* on *C*, since *D* ensures that *J* would still have occurred even if *C* had not occurred, as illustrated below:

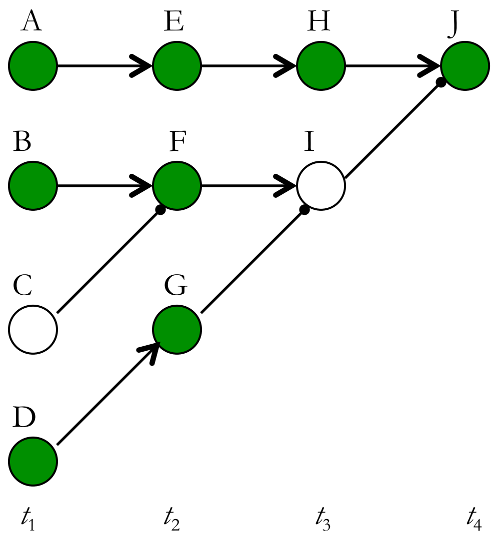


Figure 5*

Thus, we find that the two concepts account is not permissive enough: it allows no sense in which *C* is a cause of *J*, even though it seems intuitively correct to say that *C* causes *J*.[33]

2.2.3 Overview of counterexamples to the two concepts account

Figure 6 sums up the counterexamples we have considered above: the grey dots that lie within the circles of production and dependence represent counterexamples to the sufficiency of the two concepts account; the black dot outside the circles represents a counterexample to the necessity of the two concepts account.

[33] Hall (2007*b*), p. 52.

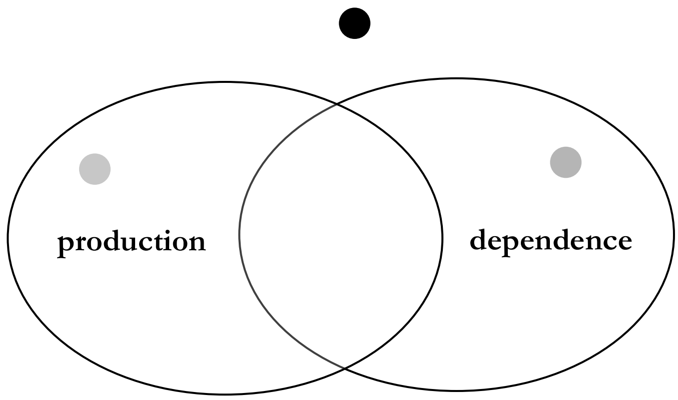


Figure 6

Given the aim of identifying the relation(s) that strike the best balance between eligibility and charity to our use of 'cause', the above considerations do not *refute* the two concepts account.[34] However, the problems do suggest that it may be worthwhile to keep searching for an account of causation that can do better in terms of satisfying the two criteria of eligibility and charity to use.

2.3 The butterfly account: causation at the intersection of process-connection and security-dependence

The crucial insight underlying Hall's two concepts account is that we can accommodate our conflicting intuitions concerning the sufficiency of counterfactual dependence, transitivity, and intrinsicness, by giving an account of causation in terms of two different relations: one relation for which counterfactual dependence is not sufficient, but which is transitive and intrinsic – namely, the relation of production; and one for which counterfactual dependence *is* sufficient, but which is neither transitive nor intrinsic – namely, the relation of dependence.

Hall implements this idea by giving very demanding conditions for production and dependence, so that there are clear cases of causation that do not satisfy one or the other of the two conditions. In particular, Hall's

[34] Cf. Hall (2007*b*), p. 52.

condition for production is so demanding that it cannot be satisfied in any case of omission-involving causation, and Hall's condition for dependence – namely, counterfactual dependence – is so demanding that it cannot be satisfied in cases of redundant causation. To accommodate our intuitive judgements, Hall must therefore take each condition on its own to be sufficient for (a kind of) causation.

However, there is a different way of implementing Hall's insight: instead of focusing on a relation of production characterised by conditions that are so demanding that production must be taken to be *sufficient* for causation, we may focus on a broader relation of production that is merely *necessary* for causation. And similarly, instead of focusing on a relation of dependence characterised by conditions that are so demanding that dependence must be taken to be *sufficient* for causation, we may focus on a broader relation of dependence that is merely *necessary* for causation. And we may then take these two relations to be *jointly sufficient* for causation.

This is exactly the strategy I adopt in my proposed account of causation. I here identify a broader relation of production, namely the binary relation of *process-connection*: c is process-connected to e. And I identify a broader relation of dependence, namely the ternary relation of *security-dependence*: e security-depends on e within possibility horizon $\mathscr{H}$. Importantly, both of these broader relations can involve omissions. And my suggestion is that these two relations are individually necessary and jointly sufficient for causation.

The new picture and its relation to Hall's account is presented in the figure below – which looks like a butterfly:

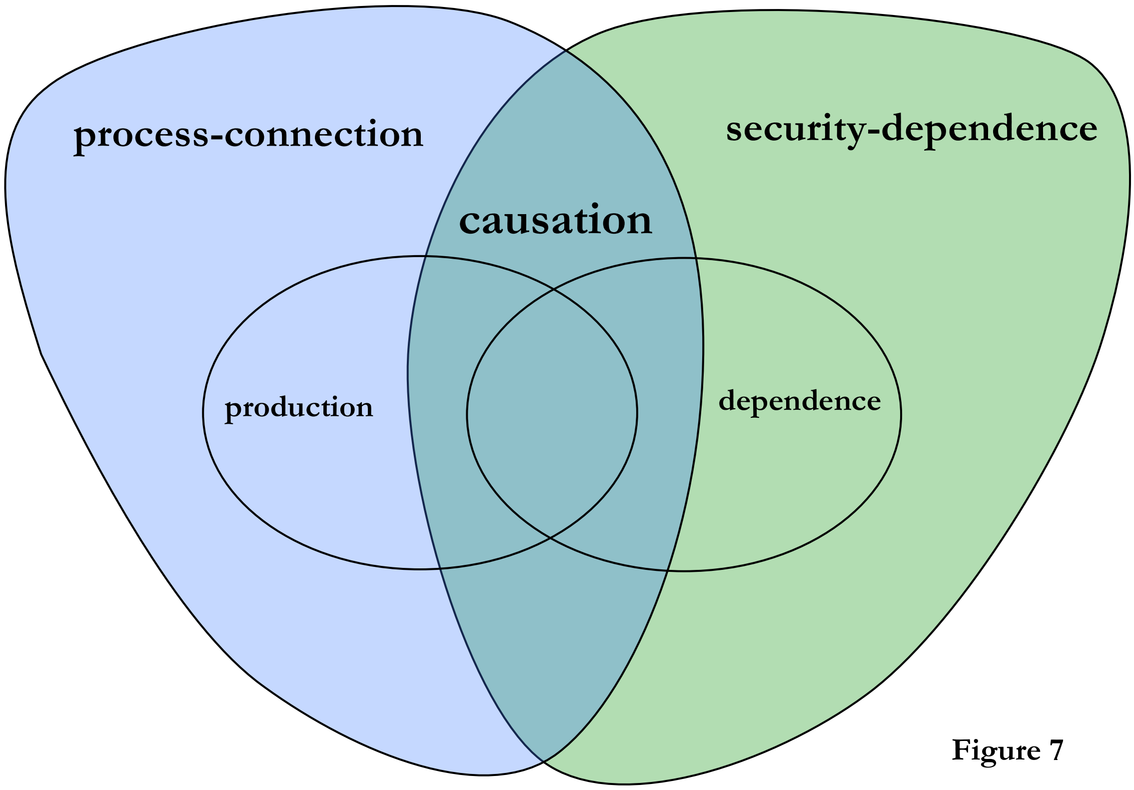


Figure 7

More precisely, my proposed account of causation may be stated as follows:

> *Causation*: *c* is a cause of *e* within a possibility horizon $\mathscr{H}$ iff
>
> a) *c* is process-connected to *e*, and
>
> b) *e* security-depends on *c* within $\mathscr{H}$.

It will, of course, require a great deal of work to set out the conditions characterising process-connection and security-dependence. Doing so will keep us occupied in the following chapters. However, we may already at this point note some crucial features of my proposed account:

My proposed account has two individually necessary and jointly sufficient conditions. The first condition – the condition of process-connection – captures the intuitive idea that a cause must be connected to its effect via a genuine process; the second condition – the condition of security-dependence – captures the intuitive idea that a cause must make a difference to its effect.

This allows me to retain the central insight of Hall's account: that we can accommodate our conflicting intuitions concerning the sufficiency of counterfactual dependence, transitivity, and intrinsicness, by recognising the dual nature of causation. Just like Hall's account, my account is made up of two relations: one relation for which counterfactual dependence is not sufficient, but which is transitive and intrinsic – namely, the relation of process-connection; and one relation for which counterfactual dependence *is* sufficient, but which is neither transitive nor intrinsic – namely, the relation of security-dependence.

By recognising the dual nature of causation in this way, my account achieves a high degree of charity to use: as we shall see in the following, my account accommodates our intuitive judgements on the counterexamples to the sufficiency of counterfactual dependence, transitivity, and intrinsicness (section 1.3 above). At the same time, however, each of the three principles applies to either process-connection or security-dependence. And as I show in Chapter 11, this allows us to recover restricted principles of sufficiency of counterfactual dependence, transitivity, and intrinsicness, where these restricted principles are still strong enough to sanction almost all of our ordinary practices of causal inference.

2.4 Preview of the following chapters

In this section, I give a brief preview of how I will develop my account of causation in the following:

Part II: Foundations

In this part, I lay the foundations for my account of causation. I first set out my conception of the laws of nature (Chapter 3), where I follow Maudlin's proposal that we should understand the fundamental laws of nature as describing 'the evolution of physical states through time'.[35] Building on this, I present a new conception of the primary causal relata as *instantaneous events* (Chapter 4). And finally, I suggest that the causal relation has a third relatum: *possibility horizons* (Chapter 5).

Part III: Process-connection

I here set out my first necessary condition for causation – namely, *process-connection*. I first define the relation of process-connection, which is a binary relation between instantaneous events (Chapter 6). I then consider the central applications of the condition. In particular, I show that the condition of process-connection allows us to successfully distinguish causation from mere correlation, differentiates between genuine causes and preempted backups in cases of redundant causation, gives a uniform treatment of ordinary and omission-involving causation, and ensures that the cause is at the right level of detail relative to its effect (Chapter 7).

Part IV: Security-dependence

I here set out my second necessary condition for causation – namely, *security-dependence*. Security-dependence is a counterfactual-based condition, and the relation of security-dependence has three relata: two instantaneous events and a possibility horizon. I begin by giving a recipe for evaluating counterfactuals within a possibility horizon, and build on that to define the relation of security-dependence (Chapter 8). In the following two chapters, I then show how the condition of security-dependence completes my account of causation, allowing it to give intuitively correct verdicts on the counterexamples to the transitivity and intrinsicness of causation (Chapter 9), and resolve the problem of profligate omissions, handle structurally isomorphic but causally different cases, and accommodate contrastive causal claims (Chapter 10).

[35] Maudlin (2007*b*), p. 12.

Part V: The biconditional

In this part, I assess my proposed account of causation. I show how my account logically entails restricted versions of each of the three intuitive principles of causal reasoning – concerning the sufficiency of counterfactual dependence, transitivity, and intrinsicness (Chapter 11). And finally, I assess how well my proposed account achieves the aim of finding the candidate meaning of 'cause' that strikes the best balance between eligibility and charity to use (Chapter 12).

Appendix

In the appendices, I offer proofs that the relation of process-connection satisfies a restricted principle of sufficiency of counterfactual dependence, that process-connection is transitive and intrinsic to a process, and that counterfactual dependence is sufficient for security-dependence (Appendix A). In addition, I give an overview over the cases discussed in the dissertation (Appendix B). These appendices are intended simply as a service to the reader, and the dissertation can be read without them.

3. Conclusion

In this chapter, I have argued that our use of 'cause' suggests that the causal relation out in the world has a dual nature: on the one hand, a cause must be connected to its effect via a genuine process; on the other hand, a cause must somehow make a difference to its effect. Furthermore, I have given a preview of how I will develop this idea: by proposing two necessary and jointly sufficient conditions for causation – the conditions of process-connection and security-dependence.

PART II

Foundations

3

Laws of nature

The laws of nature lie at the very heart of causation. In this chapter, I draw on work by Maudlin to set out a scientifically informed conception of the laws of nature.[1] I begin by presenting Maudlin's conception of the laws of nature as describing 'the evolution of physical states through time'[2] (section 1). Next, I set out in more detail how we should understand the central notion of a complete state (section 2). I then give a precise definition of the class of worlds I will be concerned with throughout this dissertation – namely, worlds governed by forwards deterministic laws that are time-translation invariant and do not permit action at a temporal distance (section 3). And finally, I set out how neuron diagrams should be understood within this framework (section 4).

1. The laws of nature as rules of temporal evolution

For the purpose of this dissertation, I adopt a minimal, widely accepted, and intuitively plausible way of thinking about the laws of nature. This way of thinking about the laws of nature has been carefully defended by Maudlin,[3] and is summed up by Hall as follows:

> 'I think of [the fundamental laws that govern what happens] on the model of physics: namely, as something like rules that specify how complete physical states of the world generate successive physical states.'[4]

[1] Maudlin (2007*b*).
[2] Maudlin (2007*b*), p. 12.
[3] Maudlin (2007*b*). Cf. Kutach (2007); Williamson (2016), 472-79; Woodward (2007).
[4] Hall (2002), p. 261.

Both Maudlin and Hall talk in terms of complete *physical* states. However, their proposed way of thinking about the laws of nature need not involve a commitment to physicalism. In the following, I therefore prefer to state the view as being concerned simply with complete states of the world, leaving it open whether such complete states can be fully characterised by final physics.

We may think of a complete state of the world as a complete state of the world at a time *t*, where the time-*t* slice of the world is defined relative to a frame of reference. The relativisation to a frame of reference is needed because, as the theory of relativity tells us, there is no absolute simultaneity, only simultaneity relative to a frame of reference.[5] In the following, however, I will leave this relativisation to a frame of reference implicit in order to avoid unnecessary repetitions.[6]

Based on this, we may now think of the fundamental laws of nature as *transition rules*: as Maudlin writes, the laws determine 'how specified states of a system will or can evolve into other states.'[7] To make this picture more precise, our next task is to get a better understanding of the crucial notion of a complete state.

2. Complete states

What are complete states? Complete states behave like properties: different worlds can be in the same complete state, and the same world can be in the same complete state at several different times.[8] Based on this, I suggest that

[5] Cf. Maudlin (2007*b*), p. 18.

[6] The relativisation to a frame of reference must, obviously, refer to an *allowed* reference system (t, x, y, z) in the sense of physics. I defer to physics on what it takes for a reference system to be allowed. For the purpose of this dissertation, I only need that allowed reference systems do exist. One simple example is the almost inertial system wherein our solar system, taken as a whole, is at rest.

[7] Maudlin (2007*b*), p. 14.

[8] Cf. Williamson (2016), p. 474.

complete states *are* properties. More precisely, I suggest that they are properties of instantaneous time-slices of worlds.[9]

To make this suggestion more precise, we need a way of thinking about properties. In the following, I will adopt Lewis's conception of properties as *classes*.[10] On this view, the property of *being red*, for example, is a class – namely, the class of all actual and merely possible red things. And a thing – a tomato, say – has the property of *being red* just in case it belongs to the class of red things.

We may similarly think of complete states as classes of instantaneous time-slices of worlds. More precisely, my suggestion is that complete states are equivalence classes based on the equivalence relation of *being a perfect duplicate of*.[11] Note that this is indeed an equivalence relation, i.e. it is reflexive, symmetric, and transitive: any instantaneous time-slice is a perfect duplicate of itself; if x is a perfect duplicate of y, then y is a perfect duplicate of x; and if x is a perfect duplicate of y, and y is a perfect duplicate of z, then x is a perfect duplicate of z. Based on this, we may define complete states as follows:

> *Complete states*: a class s of instantaneous time-slices of worlds corresponds to a complete state if and only if s is an equivalence class based on the equivalence relation of *being a perfect duplicate of*.

An instantaneous time-slice of a world is in a complete state s just in case it belongs to the corresponding equivalence class. And, by extension, a world w is

---

[9] Mathematically, we may make the idealisation of thinking about the time-width of an instantaneous time-slice as being an infinitesimal time dt.

[10] See e.g. Lewis (1986*a*), pp. 50-51. This particular conception of properties is not essential to my view. What is essential, however, is a conception of properties on which properties are *abundant*, such that for any class of things, there is a property shared by all and only the things in that class.

[11] This suggestion fits with Williamson's more general suggestion that states correspond to equivalence classes of world-time pairs under some relevant equivalence relation (Williamson (2016), p. 474).

in a complete state *s* at a time *t* just in case the instantaneous time-*t* slice of *w* belongs to the relevant equivalence class.

As we shall now see, the above definition entails three crucial results concerning complete states: that complete states are unique, intrinsic, and (as their name implies) complete. Each of these results fits our intuitive understanding of what a complete state is, and each brings out a feature of complete states that I will be assuming in the following.

First, the above definition entails the following principle of uniqueness:

> *Uniqueness*: for every world *w* and time *t*, there is exactly one complete state *s* such that *w* is in the complete state *s* at time *t*.

This result follows directly from the above definition of complete states: the equivalence relation of *being a perfect duplicate of* induces a partition on all instantaneous time-slices of worlds (this follows from the general result that every equivalence relation induces a partition). This means that each instantaneous time-slice gets to belong to exactly one equivalence class based on this relation – and thus, there is for each instantaneous time-slice exactly one complete state that this time-slice is in. By extension, we find that for every world *w* and time *t*, there is exactly one complete state *s* such that *w* is in the complete state *s* at time *t*.

Second, complete states are intrinsic properties of instantaneous time-slices. The following gives an initial gloss of how I understand intrinsicness:

> '[take] "intrinsic" to mean something like "internal" or "metaphysically independent"; intuitively, the way something is intrinsically is the way it is independent of how anything else is.'[12]

[12] Paul and Hall (2013), p. 124.

In particular, it is widely recognised that intrinsicness and perfect duplication are interdefinable. For example, Lewis defines intrinsic properties of regions as follows:

> 'We can define an *intrinsic* property of a region as one such that, whenever two possible regions are perfect duplicates, the property belongs to both or neither.'[13]

We may similarly define an intrinsic property of an instantaneous time-slice as one such that, whenever two instantaneous time-slices are perfect duplicates, the property belongs to both or neither. Based on this, it follows immediately that complete states are intrinsic properties of instantaneous time-slices:

> *Intrinsicness*: complete states are intrinsic properties of instantaneous time-slices.

Third, the interdefinability between intrinsicness and perfect duplication entails that perfect duplicates share *all* of their intrinsic properties. Thus, we find that complete states are indeed complete, in the sense that specifying the complete state of an instantaneous time-slice is tantamount to specifying *all* the intrinsic properties of that time-slice: complete states are so rich and detailed that each complete state entails every intrinsic property of the instantaneous time-slices that instantiate it – this is a direct result of the fact that only perfect duplicates can be in the same complete state. We may state this principle as follows:

> *Completeness*: two instantaneous time-slices are in the same complete state only if they share *all* their intrinsic properties.

---

[13] Lewis (1986*g*), p. 263. Intrinsicness and perfect duplication belong to a small circle of interdefinable notions. It is much harder to give a definition of intrinsicness that breaks out of this circle. For discussion, see e.g. Langton and Lewis (1998); Marshall and Parsons (2001); Sider (2001) and (2003); and Weatherson (2001).

We have now seen how the three principles – concerning uniqueness, intrinsicness, and completeness – flow from my definition of complete states as equivalence classes of instantaneous time-slices of worlds, based on the equivalence relation of *being a perfect duplicate of*. Thus, the three principles all flow from a single coherent picture. For my purposes, however, all that matters is the general understanding of complete states as properties that satisfy the three principles. Any other way of arriving at such an understanding of complete states is just as good for my purposes. Indeed, you may if you like take complete states themselves as primitive, and impose the principles of uniqueness, intrinsicness, and completeness as postulates. All that matters is that the three principles are satisfied one way or another.

Based on this understanding of complete states, I will now give the definitions – of determinism, time-translation invariance, etc. – that allow us to characterise the worlds I will be concerned with in this dissertation.

3. Determinism, time-translation invariance, and no action at a temporal distance

As we have seen above, Maudlin characterises the laws of nature simply as transition rules describing 'how specified states of a system will or can evolve into other states.'[14] As Maudlin shows, this way of thinking about the laws of nature may be applied to both deterministic and indeterministic laws: when the laws are deterministic, they specify how specified states of a system *will* evolve into other states; when the laws are indeterministic, they specify how specified states of a system *can* evolve into other states (and, typically, such indeterministic laws also specify a probability distribution over the different states that a system may evolve into).[15]

In the following, I will focus on developing an account of causation that applies to worlds with forwards deterministic laws that satisfy certain further

[14] Maudlin (2007*b*), p. 14.

[15] Maudlin (2007*b*), p. 14.

requirements. Following standard practice, we may give the following simple definition of what it takes for a set of laws to be forwards deterministic:

> *Forwards deterministic laws*: a set of laws L is *forwards deterministic* iff for any two nomologically possible worlds *w* and *w´*, it is the case that if the history of *w* and *w´* is exactly the same up until time *t*, then the future history of *w* is exactly the same as the future history of *w´*.[16]

In the following, I only consider worlds with forwards deterministic laws. In addition, I impose further restrictions on the worlds I consider. First, I follow Paul and Hall in considering only worlds where the laws do not permit action at a temporal distance.[17] Given forwards determinism, the prohibition against action at a temporal distance may be stated as follows:

> *No action at a temporal distance*: for any two nomologically possible worlds, *w* and *w´*, if the complete state of *w* at *t* is the same as the complete state of *w´* at *t*, then the future history of *w* is exactly the same as the future history of *w´*.[18]

As Paul and Hall note, this prohibition against action at a temporal distance amounts to the requirement that 'the present state of the world renders facts about the past irrelevant to what happens in the future'.[19]

---

[16] See Paul and Hall (2013), p. 8 footnote 4. Cf. Lewis (1986*d*), p. 162. For the definition to successfully capture what we mean by saying that a given set of laws is forwards deterministic, we need to home in on the right understanding of what is included in the history of a world up until time *t*. The conception of complete states proposed above, according to which complete states are unique, intrinsic, and complete, allows us to do just that: based on this conception of complete states, we may now say that two worlds *w* and *w´* have the same history up until time *t* iff it is the case that for each time up until time *t*, *w* and *w´* are in exactly the same complete state at that time.

[17] Paul and Hall (2013), p. 8.

[18] Paul and Hall (2013), p. 8 footnote 4. Note that this statement presupposes the above characterisation of complete states.

[19] Paul and Hall (2013), p. 8 footnote 4.

Finally, I impose the restriction that the laws must be time-translation invariant. Given forwards-determinism and the prohibition against action at a temporal distance, time-translation invariance may be defined as follows:

> *Time-translation invariance*: if a nomologically possible world w is in the complete state s_1 at time t, and in the complete state s_2 at time $t+dt$, then for any nomologically possible world w' and any time t', it is the case that if w' is in the complete state s_1 at time t', then w' is in the complete state s_2 at time $t'+dt$.[20]

My hope is that my account of causation may eventually be generalised to cover all metaphysically possible worlds. For now, however, my aim is simply to develop the account so that it holds within worlds that do satisfy the requirement of being governed by forwards deterministic laws that are time-translation invariant and do not permit action at a temporal distance.

4. Neuron diagrams and complete states

As I have mentioned already (cf. Chapter 2 section 1.1), I will often rely on neuron diagrams to represent causal structures. In this section, I show how we may understand neuron diagrams in a way that fits nicely with my characterisation of the laws of nature as rules specifying the temporal evolution of complete states. Consider, for example, our standard case of early preemption illustrated below:

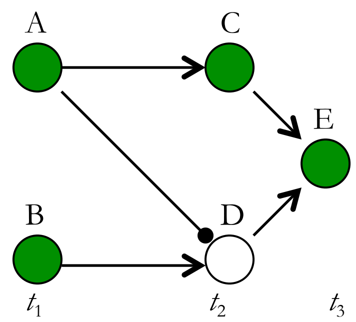


Figure 1

[20] Note that this definition presupposes the above characterisation of complete states: without this, the definition would be either too easy or too hard to satisfy.

Figure 1 gives a partial representation of the history of a very simple possible world. Treating the world of the example as actual, let us denote this world by '@'.

We may think of a neuron world as containing a physical structure of neurons connected by stimulatory and inhibitory channels. This physical structure of neurons, together with the stimulatory and inhibitory channels between them, is present *at all times*. For example, the neuron world @ contains at all times exactly five neurons **A**, **B**, **C**, **D**, and **E**, connected by stimulatory and inhibitory channels, as illustrated below:

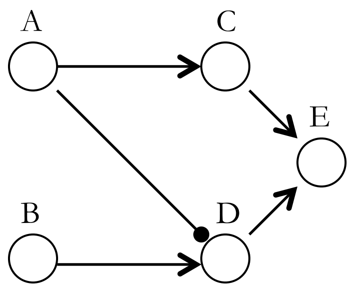


Figure 2

Furthermore, I will treat this physical structure of neurons, together with the stimulatory and inhibitory channels between them, as being nomologically necessary.[21] From the standpoint of @, for example, the structure of neurons represented in Figure 2 is nomologically necessary: every nomologically possible world contains at all times exactly five neurons, connected by stimulatory and inhibitory channels as illustrated above.

We may specify the complete state of @ at any given time *t* by specifying, for each of the five neurons **A**, **B**, **C**, **D**, and **E**, whether it does or

[21] It might seem a bit odd to treat the existence of certain neurons, together with the stimulatory and inhibitory channels between them, as nomologically necessary. I do so for the sake of simplicity: as we shall see, my account of causation requires us to consider all nomologically possible worlds. Treating the physical structure of neurons, together with the stimulatory and inhibitory channels between them, as nomologically necessary, makes this task much more manageable.

does not fire at time t. The pattern of stimulatory and inhibitory channels, together with the general neuron laws of Chapter 2 section 1.1, now yields laws for the temporal evolution of such complete states. And it is easily verified that these laws are forwards deterministic, time-translation invariant, and obey the prohibition against action at a temporal distance.

Based on the physical structure of neurons illustrated in Figure 2, we can distinguish $2^5 = 32$ different possible complete states, corresponding to the different combinations of whether **A**, **B**, **C**, **D**, and **E** fire or fail to fire: there is a complete state such that **A**, **B**, **C**, **D**, and **E** all fail to fire; there is a complete state such that **A** fires, but **B**, **C**, **D**, and **E** fail to fire, etc. A nomologically possible world may start out in any of the 32 different complete states. Since the laws are forwards deterministic, specifying the initial complete state of a world is sufficient to specify the complete history of the world.

By contrast, our standard neuron diagram gives only a partial representation of the history of a possible world. To see this, consider once again Figure 1. First, the neuron diagram in Figure 1 only explicitly shows what happens in @ at three times – namely, t_1, t_2, and t_3. Second, the neuron diagram only gives a partial specification of the complete state of @ at each of these three times. Concerning time t_1, it tells us that **A** and **B** fire, but it remains silent on what **C**, **D**, and **E** are doing. Concerning time t_2, it tells us that **C** fires and **D** does not, but it remains silent on what **A**, **B**, and **E** are doing. And concerning time t_3, it tells us that **E** fires, but remains silent on what **A**, **B**, **C**, and **D** are doing.

What shall we make of these partial representations, where, for example, the neuron diagram in Figure 1 only specifies the state of the two neurons **A** and **B** at time t_1? The answer is simple: we may see this specification concerning time t_1 as giving a *class* of complete states, characterised by the condition that both **A** and **B** fire. This class contains $2^3 = 8$ complete states. And the neuron diagram tells us that @ is in *one* of these complete states at time t_1, but it does not specify which one. Similarly, the information given by the neuron diagram concerning time t_2 specifies a class of complete states characterised by the condition that **C** fires and **D** does not. This equivalence class once again contains $2^3 = 8$ complete states, and the neuron diagram tells us that @ is in *one* of these complete states at time t_2, but does not specify which one. And finally, the information given by the neuron diagram concerning time t_3 specifies a class of complete states characterised by the condition that **E** fires. This equivalence class contains $2^4 = 16$ complete states, and the neuron diagram tells us that @ is in *one* of these complete states, but does not specify which one.

Thus, the neuron diagram in Figure 1 tells the history of @ in the following way: @ starts out at time t_1 in *some* complete state such that **A** and **B** both fire; this evolves so that @ is at time t_2 in *some* complete state such that **C** fires and **D** does not fire; and this finally evolves so that @ is at time t_3 in *some* complete state such that **E** fires.

Given the neuron laws, the partial specifications that the neuron diagram provides of what happens at t_2 and t_3 are strictly speaking redundant: it is easily verified that *any* nomologically possible world that starts out at time t_1 in a complete state such that **A** and **B** both fire will subsequently evolve so that its complete state at t_2 is such that **C** fires and **D** does not, and so that its complete state at t_3 is such that **E** fires. Showing this in the neuron diagram is simply a matter of convenience.

It is worth noting that this redundancy is a general feature of neuron diagrams: all complete states that agree on what the neurons explicitly shown at a time t_i are doing, also agree on what the neurons explicitly shown at time t_{i+1} are doing. If we are only interested in the behaviour of neurons at the times when their behaviour is explicitly shown in the neuron diagram – for example, if we are only interested in the behaviour of **A** and **B** at time t_1, the behaviour of **C** and **D** at time t_2, and the behaviour of **E** at time t_3 – this allows us to treat the neuron diagram *as if* it gives the complete history of an even simpler system evolving in three stages:

In the first stage, at time t_1, this system contains exactly two neurons, **A** and **B**, and we can distinguish just $2^2 = 4$ different complete states at this time, corresponding to the different combinations of whether **A** and **B** fire or fail to fire. In the second stage, at time t_2, the system contains exactly two neurons **C** and **D**, and we can again distinguish just $2^2 = 4$ different complete states at this time, corresponding to the different combinations of whether **C** and **D** fire. And in the third stage, at time t_3, the system contains only neuron **E**, and we can distinguish only two complete states – the complete state such that **E** fires, and the complete state such that **E** does not fire.

This simplification is entirely harmless, and in the following I will often, for the sake of simplicity, speak *as if* it is correct. However, one should keep in mind that the correct interpretation of the neuron diagrams we are working with is in fact the more complex interpretation set out above.

5. Conclusion

The conception of the laws of nature I have adopted here – where the laws of nature are understood simply as transition rules from one complete state to another – is all I need to develop my proposed account of causation. This allows me to remain neutral on some of the central questions concerning how we should understand the laws of nature.

First, I remain neutral on the question of whether the fundamental laws of nature simply are the individual transition rules, or whether the fundamental laws are rather generalisations from which these specific transition rules follow. Maudlin's view on this question seems to be that the fundamental laws are generalisations: he considers a case where we know all the specific transition rules, but where this still 'does not settle what the laws are'.[22] For my purposes, however, nothing hangs on this: my account is equally compatible with both answers to the question.

[22] Maudlin (2004), p. 426.

Second, I remain neutral on the question of whether we should adopt a Humean conception of the laws of nature. This is, in Hall's words, the question 'whether facts about the fundamental laws themselves reduce to facts about the totality of physical states the world occupies.'[23] Maudlin's position is distinctly non-Humean: he argues that status as a fundamental law is primitive and suggests a picture on which the laws of nature are in some deep sense *productive*, generating each successive complete state of the world.[24] Once again, however, nothing in my proposed account of causation hangs on this: my account is equally compatible with a Humean picture, such as the one developed by David Lewis.[25]

By adopting this minimal and widely accepted conception of the laws of nature, I hope to ensure that my account of causation is compatible with as wide a range of positions on the laws of nature as possible. In the following chapter, we shall now see how even this minimal conception of the laws of nature is still enough to guide us towards a new picture of the causal relata.

[23] Hall (2007*b*), p. 3 footnote 6.

[24] Maudlin (2007*b*), p. 15.

[25] See e.g. Lewis (1986*b*), pp. xi-xii.

4

The primary causal relata: instantaneous events

Causation is a relation. What are its relata?

My suggestion is that causation is a ternary relation: *c* causes *e* within a possibility horizon $\mathscr{H}$. The primary causal relata are the cause *c* and the effect *e*. In this chapter, I present my proposed conception of the primary causal relata. In Chapter 5, I then introduce the notion of a possibility horizon.

There is a *prima facie* tension between our ordinary understanding of the primary causal relata and the scientifically informed conception of the laws of nature set out in Chapter 3. On our ordinary understanding, the primary causal relata are events – where these events are understood to be *local* and *temporally extended*. Examples include Suzy's throwing a rock, the shattering of a window, an icy cavern collapsing in a Greenland glacier, etc.

On the scientifically informed conception set out in Chapter 3, on the other hand, we may think of the laws of nature as transition rules specifying how complete states evolve through time. These complete states are properties of instantaneous time-slices of worlds, and thus their concrete instantiations are *global* and *instantaneous*. Thus, there is a tension between our ordinary conception of the primary causal relata and a scientifically informed understanding of the laws of nature.[1]

To bring out this tension more clearly, let us consider the example of Suzy throwing a rock towards a window. On our ordinary understanding of the primary causal relata, we think of Suzy's throw as a local and temporally extended event: it is located just where Suzy is, and it is temporally extended –

[1] See e.g. Kutach (2007) and Woodward (2007).

lasting (with some fuzziness at the boundaries) from the moment when Suzy pulls back her arm to the moment when the rock leaves her hand. To determine whether Suzy's throw causes the window-shattering, it now seems reasonable to look towards the laws of nature. However, the laws of nature are not concerned with local and temporally extended events such as Suzy's throw. Rather, they are concerned with complete states – such as the complete state the world is in at the time *t* when the rock is just leaving Suzy's hand. This complete state is global – it includes everything that is happening at time *t*: the rock is leaving Suzy's hand with a certain velocity, a blackbird is singing from a tree nearby, a dust storm is raging on a planet in a galaxy far, far away, etc. And it is instantaneous – it concerns just the time *t*, not the extended period of time during which Suzy throws.

This mismatch creates a puzzle about how the primary causal relata connect up with the laws of nature. My aim in this chapter is to propose a simple, new conception of the causal relata that resolves this puzzle by establishing a clear connection between the primary causal relata and the complete states that feature in the laws of nature, while allowing us to recover the important features of our ordinary understanding of the causal relata.

My suggestion is that the primary causal relata are *instantaneous events*. I give a precise definition of instantaneous events in section 2. In particular, the primary causal relata are *actual instantaneous events*, where an actual instantaneous event is an instantaneous event that occurs in the actual world.

In brief, an actual instantaneous event has two elements: its *realization* and its *modal profile*. The realization consists in the actual world being in a particular complete state at a particular time. The modal profile specifies what is essential and what is merely accidental to the occurrence of the event.

Consider, for example, Suzy throwing a rock. We may then characterise an actual instantaneous event *e* based on Suzy's throw as follows: the realization of *e* consists in the actual world being in a particular complete state at a time *t* during Suzy's throw – say, the complete state such that the rock is leaving Suzy's hand with a certain velocity, a blackbird is singing from a tree nearby, a dust storm is raging on a far-away planet, etc. And the modal profile of *e* is such that it is essential to the occurrence of *e* that Suzy throws, whereas it is only accidental to the occurrence of *e* that a blackbird is singing from a tree nearby, or that a dust storm is raging on a far-way planet, etc.

By paying attention to the modal profiles of instantaneous events, we can recover the ordinary sense in which the causal relata are local: although an instantaneous event *e* is realized by the actual world's being in a complete state *s*, which is of course global, the features of *s* that are essential to the occurrence of *e* may be exhibited locally. In the case of Suzy's throw, for example, the features that are essential to the occurrence of this event are exhibited locally – namely, just where Suzy is when she throws. Furthermore, we can capture ordinary causal claims relating temporally extended events by understanding such temporally extended events as being built up from a sequence of instantaneous events. For example, the temporally extended event of Suzy's throw is built up from a sequence of many instantaneous throwing events.

The result of this is a conception of the primary causal relata that accommodates our ordinary understanding, while fitting perfectly with our scientifically informed understanding of the laws of nature. It should perhaps not be surprising that such a conception of the causal relata carries additional benefits: as we shall see, it allows us to resolve the long-standing puzzles – concerning omission-involving causation, causal differences, extensionality, and transitivity – that are standardly used to test accounts of the causal relata.[2] Furthermore, the details of this conception of the causal relata turn out to be crucial in allowing us to define the relation of process-connection, as we shall see in Chapter 6.

In the following, I begin with a few terminological remarks (section 1), and then develop my proposed conception of the primary causal relata as instantaneous events (section 2). I then show how my proposed conception of

[2] For an overview, see Schaffer (2016).

the causal relata allows us to recover our usual understanding of the causal relata as local and temporally extended events (section 3). In the subsequent three sections, I discuss the central applications of this proposal: I show how it allows us to accommodate omission-involving causation (section 4), how it allows us to respect causal differences and preserve extensionality (section 5), and how it allows us to resist a particular kind of counterexample to the transitivity of causation (section 6). Finally, I discuss restrictions on the domain of instantaneous events (section 7).

1. A note on terminology

Causal claims take a variety of different forms – they may feature events, facts, agents, etc., in the role of cause and effect.[3] It is reasonable to aim for an account of causation that can accommodate causal claims of all these different forms. To achieve this, however, we need not allow for primary causal relata of all the different categories – events, facts, agents, etc. – suggested by the surface form of our causal claims. Instead, we may use the widespread strategy of taking claims of all these different forms to be true (or false) in virtue of relations between a unique kind of primary causal relata – what we might call the *basic* causal relata. This is, for example, the strategy that Paul and Hall adopt, taking the basic causal relata to be events (at least for the purposes of most of their subsequent discussion):

> 'Causation seems, at least in the first instance, to relate *events*: while Suzy might cause a window to break, she does so only in virtue of the way she is involved in an event – her throwing of a rock, say – that causes the breaking.'[4]

In the following, I will adopt the same strategy of taking causal claims to be true (or false) in virtue of relations between the basic causal relata. The question I aim to answer in this chapter is: what are the basic causal relata?

[3] Schaffer (2016), pp. 5-6.
[4] Paul and Hall (2013), p. 4.

And my proposed answer is that the basic causal relata are instantaneous events. In the following, I shall simply refer to the basic causal relata as 'the causal relata'.

My account of instantaneous events is intended to capture a clean, technical notion that is specifically tailored to the role of causal relatum. Occasionally, I will need to compare and contrast this with our ordinary, everyday notion of events – Suzy's throwing a rock, the shattering of a window, etc., are paradigm examples. To keep a clear distinction between these two notions, I will in the following always refer to ordinary events as 'ordinary events', whereas I may sometimes for convenience use 'event' as shorthand for 'instantaneous event'.

2. Instantaneous events

In this section, I set out my proposed conception of the causal relata as instantaneous events (section 2.1). I then discuss how one instantaneous event may be a more fragile version of another (section 2.2), and set out conventions for talking about neuron firing events (section 2.3).

2.1 Defining instantaneous events

I suggest that the causal relata are *instantaneous events*.

We may characterise an actual instantaneous event by specifying two elements: its *realization* in the actual world and its *modal profile*. The realization specifies how the instantaneous event is realized in the actual world and may be characterised by a triple (@, t, s), where @ is the actual world, t is a time, and s is the complete state of @ at t.[5] The modal profile specifies what is

[5] There is an intuitive way of thinking about ordinary events according to which they are property exemplifications: an ordinary event consists in an object exemplifying a property at a time (see Kim (1973) and (1976)). For example, the shattering of a window consists in an object, namely the window, exemplifying a property, namely shattering, at a time. The realization (@, t, s) of an actual instantaneous event is such a property exemplification, although at a global scale: it simply consists in an object, namely the actual world @, exemplifying a property, namely a complete state s, at the relevant time t.

essential to the occurrence of the event. This may be characterised by a pair (I, C), where I is a closed interval that includes time t, and C is a class of complete states that includes s.

Thus, an actual instantaneous event is characterised by a quintuple (@, t, s, I, C) that specifies first the event's actual realization (@, t, s), and next its modal profile (I, C). By contrast, a merely possible instantaneous event, i.e. an instantaneous event that does not occur in @, may be characterised simply by its modal profile (I, C), or – in cases where we want to specify that it occurs in a world w by virtue of a realization (w, t, s) – by a quintuple (w, t, s, I, C).

At any given time t, the actual world is in exactly one complete state s. It follows from this that all actual instantaneous events that occur at a given time t have the same realization – namely, the realization characterised by the triple (@, t, s), where s is the complete state of @ at t. This realization is truly global: it includes everything that is going on at time t – my drinking a cup of tea, a bumblebee finding a clover blossom, a speck of sand being caught by the wind in the Sahara, an iceberg larger than the state of Delaware breaking off from an ice shelf in Antarctica, an atom undergoing radioactive decay somewhere in the Andromeda galaxy, etc.[6]

Intuitively, however, the causes and effects of my drinking a cup of tea are different from the causes and effects of the bumblebee finding the clover

[6] It seems plausible that some of the predicates that feature in these descriptions – such as 'drinking a cup of tea', 'being a bumblebee', etc. – take time to be satisfied. For example, one might suggest that I satisfy the predicate 'drinking a cup of tea' only by spending some time taking a sip from my cup of tea once in a while. This means that it cannot be intrinsic to an instantaneous time-slice that I am drinking a cup of tea. Rather, an instantaneous time-slice is such that I am drinking a cup of tea in virtue of two components: the first component concerns the intrinsic properties of the instantaneous time-slice – it has to contain an instantaneous time-slice of me and an instantaneous time-slice of a cup of tea within suitable distance from each other; the second component concerns the relational properties of the instantaneous time-slice – it has to be suitably related to earlier and later instantaneous time-slices, where I continue my tea-drinking (this suggestion is entirely parallel to Hawley's treatment of lingering and historical properties in Hawley (2004), pp. 53-54). By extension, we may then say that a complete state is such that I am drinking a cup of tea just in case this complete state shares all the intrinsic properties of an instantaneous time-slice such that I am drinking a cup of tea.

blossom, which are in turn different from the causes and effects of the speck of sand being caught by the wind in the Sahara, etc. My suggestion is that while all these events have the same realization, they are distinct because they have different modal profiles: events are not merely individuated based on their realization; it also matters for the identity of an event what is essential to its occurrence, and what is merely accidental.[7]

This allows us to capture the intuition that my drinking a cup of tea and the bumblebee finding the clover blossom are distinct causal relata, although they have the same actual realization: it is essential to the event of my drinking a cup of tea that the complete state of the world is such that I am drinking a cup of tea, whereas it is only accidental that it is such that the bumblebee finds the clover blossom. By contrast, it is essential to the event of the bumblebee finding the clover blossom that the complete state of the world is such that the bumblebee finds the clover blossom, while it is only accidental that it is such that I am drinking a cup of tea.

This is why we need the modal profile (I, C) in our characterisation of an instantaneous event. To see how this works, note that we may specify what is essential and what is merely accidental to the occurrence of an event e along two dimensions: first, we may specify which properties of the time t are essential to the occurrence of e – this is specified by the interval I. And second, we may specify which properties of the complete state s are essential to the occurrence of e – this is specified by the class of complete states C. In the following, I will explain both of these points in more detail. In explaining both of these points, I shall adopt a Lewisian conception of properties, according to which properties correspond to classes.[8]

---

[7] The suggestion that events should be individuated based on their modal profiles is found e.g. in Lewis (1986*g*) and Yablo (1992*a*) and (1992*b*). For recent discussions, see McDonnell (forthcoming) and Kaiserman (forthcoming).

[8] See e.g. Lewis (1986*a*), pp. 50-51. My proposal need not be tied to this particular view of properties: it is compatible with other views, provided that properties are treated as *abundant*, so that for every class, there is a property had by all and only entities included in the class.

On a Lewisian conception of properties, properties of times correspond to classes of times, where a time t has a given property just in case it belongs to the corresponding class. There is a clear distinction to be drawn between classes of times containing scattered times and classes containing exactly those times that together make up a closed interval. In the following, I assume that only those properties of times that correspond to closed intervals, such as the closed interval $[t_1, t_2]$, can be essential to the occurrence of events. We may thus specify which properties of a time t are essential to the occurrence of an event e by specifying the smallest closed interval I, such that it is essential to the occurrence of e that t belongs to I.

We may think of properties of complete states in a similar way: properties of complete states are classes of complete states, where a complete state s has the property corresponding to class C just in case s belongs to C. We may now capture which properties of a complete state s are essential to the occurrence of an event e by specifying the smallest class C of complete states, such that it is essential to the occurrence of e that s belongs to C.

So far, I have set out how instantaneous events may be characterised. However, I have not said what they *are*. My suggestion is that an instantaneous event is a *class* of realizations, where an event e occurs in a world w just in case e has a realization in w.[9] A realization simply consists in a world being in a particular complete state at a particular time. Thus, realizations may be individuated as follows:

> *Individuating realizations*: a realization (w, t, s) is identical to a realization (w^*, t^*, s^*) iff $w = w^*$, $t = t^*$, and $s = s^*$.

Furthermore, a realization (w, t, s) exists just in case world w is in the complete state s at time t:

[9] This proposal is structurally similar to Lewis's proposal that events are classes of space-time regions, where an event occurs in a world w just in case w contains a space-time region that belongs to the relevant class. See Lewis (1986*g*).

> *Realization*: a realization characterised by the triple (w, t, s) exists iff world w is in the complete state s at time t.

An instantaneous event e consists in a class of such realizations. This immediately yields the following principle of individuation for instantaneous events:

> *Individuation of instantaneous events*: an instantaneous event e is identical to an instantaneous event e^* iff e and e^* have the same realizations.

Importantly, however, a class of realizations has to satisfy certain constraints in order to correspond to an instantaneous event. First, a class of realizations corresponds to an instantaneous event only if it contains at most one realization in each world. This requirement ensures that instantaneous events are *non-repeatable*, so that an instantaneous event e occurs at most once in any given world.[10] Second, a class of realizations corresponds to an instantaneous event only if it is adequately characterised by a modal profile (I, C). We may sum up these requirements as follows:

> *Instantaneous event*: a class c of realizations corresponds to an *instantaneous event* iff
> a) c contains at most one realization in each world, and
> b) there is a modal profile (I, C), such that c contains a realization in a world w iff there is a realization (w, t, s), such that t belongs to I, and s belongs to C.

It should be clear from this that a modal profile (I, C) does not *uniquely* characterise an instantaneous event. Suppose, for example, that the same world w features in two different realizations (w, t_1, s_1) and (w, t_2, s_2), that both satisfy the same modal profile (I, C). In that case, there is one instantaneous event that is characterised by (I, C) and contains (w, t_1, s_1), but not (w, t_2, s_2). And

[10] Cf. Lewis (1986*g*), p. 243.

there is a different instantaneous event that that is also characterised by (I, C) and contains (w, t_2, s_2), but not (w, t_1, s_1).

Thus, specifying a modal profile does not uniquely characterise an instantaneous event: to uniquely characterise an instantaneous event, we should list every realization belonging to the relevant class. I take a step towards this by including the actual realization in my specification of actual instantaneous events: this resolves the indeterminacy in cases where there are two or more realizations in the actual world that all satisfy a particular modal profile. Beyond this, however, any indeterminacy that arises from merely specifying a modal profile is harmless.

Finally, we may say that an actual instantaneous event e occurs at time t just in case e has a realization at t:

> *Occurrence*: an actual instantaneous event $e = (@, t, s, I, C)$ occurs at time t^* iff the time t that features in its realization in @ is such that $t = t^*$.

Note that this last principle is formulated from the standpoint of the actual world @. However, it is in fact fully general, since any world w can play the role of our actual world.[11] To arrive at a general statement that applies to an arbitrary world w, one must simply replace '@' with 'w' and add the needed relativisations. For example, the generalised statement of *Occurrence* is the following:

> *Occurrence – generalised*: an instantaneous event $e = (w, t, s, I, C)$ occurs at a time t^* in world w iff the time t that features in its realization in w is such that $t = t^*$.

In the following, I will be giving all definitions from the standpoint of the actual world, since it makes the definitions simpler with no loss of generality.

[11] As Lewis writes: 'an arbitrary world w can play the role of our actual world. In speaking of our actual world without knowing just which world is ours, I am in effect generalizing over all worlds.' (Lewis (1986*d*), p. 163.)

It follows immediately from the above characterisation of instantaneous events that there is a great abundance of actual instantaneous events:

Plenitude: there exists an actual instantaneous event e = (@, t, s, I, C) iff

a) @ is in the complete state s at time t,

b) t belongs to I, and

c) s belongs to C.

It also immediately follows that at any time t, all actual instantaneous events that occur at time t have the same realization:

Same realization: an actual instantaneous event e = (@, t, s, I, C) has the *same realization* as an actual instantaneous event e^* = (@, t^*, s^*, I^*, C^*) iff $t = t^*$.

Finally, my characterisation of instantaneous events entails the following principle, showing that the modal profile (I, C) of an instantaneous event really does capture its essential properties:

Modal profile: an instantaneous event with the modal profile (I, C) occurs in a world w iff there is a time t and a complete state s such that

a) w is in the complete state s at time t,

b) t belongs to I, and

c) s belongs to C.

In the following section, we shall now see how instantaneous events may stand in interesting logical relations to each other.

2.2 Fragility

As we have just seen, all the actual instantaneous events that occur at a given time have the same realization. The difference between them is a difference in essence. To talk about this more precisely, I will rely on Lewis's concept of fragility:

> 'Call an event *fragile* if, or to the extent that, it could not have occurred at a different time, or in a different manner. A fragile event has a rich essence; it has stringent conditions of occurrence.'[12]

Following standard usage, I will say that a less fragile event is more *robust*.

Based on this concept of fragility, we may now characterise a relation that can hold between actual instantaneous events occurring at the same time: in some cases, one instantaneous event is a *more fragile version* of another, in the sense that the first event has a richer essence than the second. To capture this, the following notation will be useful:

> For any event e, $w(e)$ is the set of possible worlds in which e occurs.[13]

With this notation in place, we may now define what it takes for one actual instantaneous event to be a more fragile version of another:[14]

> *More fragile version – definition 1*: an event $e^+ = (@, t^+, s^+, I^+, C^+)$ is a *more fragile version* of an event $e = (@, t, s, I, C)$ iff
>
> a) e^+ and e have the same realization in @, and
>
> b) $w(e^+) \subseteq w(e)$.

[12] Lewis (1986*e*), p. 196.

[13] Note that this definition requires us to quantify over absolutely all metaphysically possible worlds, with no restrictions on our domain of quantification.

[14] It is straightforward to generalise this definition, yielding a definition of what it is for an event e^+ to be a more fragile version of an event e from the standpoint of any world w. For the sake of simplicity, however, I only define the relation from the standpoint of the actual world, since this is all we need in the following.

Note that this definition includes the limit case where e^+ occurs in just the same worlds as e. When an event e^+ is a more fragile version of an event e, and $w(e^+)$ is a *proper subset* of $w(e)$, we may say that e^+ is a *strictly* more fragile version of e.

It is worth noting that the above definition is equivalent to the following:

> *More fragile version – definition 2*: an event $e^+ = (@, t^+, s^+, I^+, C^+)$ is a *more fragile version* of an event $e = (@, t, s, I, C)$ iff
> a) e^+ and e have the same realization in @,
> b) $I^+ \subseteq I$, and
> c) $C^+ \subseteq C$.

With this understanding of fragility, we can now distinguish different *ways* in which one instantaneous event may be a more fragile version of another. In particular, an instantaneous event e^+ may be a more fragile version of an instantaneous event e by having more stringent conditions concerning its *time* of occurrence, while having exactly the same conditions concerning the relevant complete state. In that case, we may say that e^+ is a more *temporally* fragile version of e, defined as follows:

> *More temporally fragile version*: an event $e^+ = (@, t^+, s^+, I^+, C^+)$ is a *more temporally fragile version* of an event $e = (@, t, s, I, C)$ iff
> a) e^+ and e have the same realization in @,
> b) $I^+ \subseteq I$, and
> c) $C^+ = C$.

This notion of a more temporally fragile version will play a crucial role in my treatment of late preemption (see Chapter 6 section 3).

2.3 Neuron firing events

I rely on neuron diagrams to illustrate many of the cases I discuss throughout this dissertation (cf. Chapter 2 section 1.1 and Chapter 3 section 4). It will therefore be useful to have conventions for talking about instantaneous events based on neuron firings. In this section, I set out these conventions.

Consider, for example, the neuron diagram below:

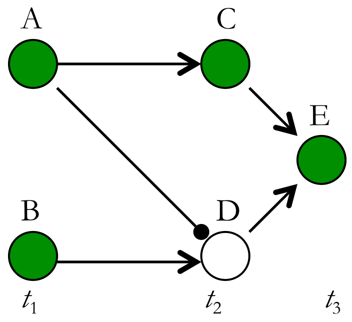


Figure 1

At each of the times t_1, t_2, and t_3, there is a great multitude of instantaneous events occurring in the neuron world illustrated here. All of the instantaneous events that occur at time t_1 have the same realization. Treating the world of the example as actual, this realization is given by the triple (@, s, t_1), where s is the complete state of @ at time t_1. And similarly in the case of the events occurring at t_2 and t_3. However, each of the different instantaneous events occurring at a given time is characterised by a different modal profile (I, C).

In the following, we only need to be concerned with a very small subset of the events that occur at any given time. First, we only need to be concerned with events that are maximally temporally fragile, so that for events that occur at time t_i, it is essential to their occurrence that they occur within the interval [t_i, $t_i + dt$].[15] Second, we only need to be concerned with events based on the firing or failure to fire of a single neuron. And third, we only need to be concerned with the behaviour of a neuron at the time when it is explicitly shown in the neuron diagram. Thus, we only need to be concerned with the behaviour of

[15] As mentioned in Chapter 3, we may make the idealisation of thinking about the time-width of an instantaneous time-slice as being an infinitesimal time dt. Using this idealisation, we shall refer to the instant t_i as the closed interval [t_i, t_i+dt], and *vice versa*.

neuron **A** and **B** at time t_1, the behaviour of neuron **C** and **D** at time t_2, and the behaviour of neuron **E** at time t_3.

To have a convenient way of naming these events, I adopt the following convention where the world of the example is treated as actual:

> For any neuron **A**, such that **A**'s firing at time t is explicitly shown in the neuron diagram, A is the instantaneous event characterised by $(@, t, s, I_A, C_A)$, where
> s = the complete state of @ at t,
> I_A = the closed interval $[t, t + dt]$, and
> C_A = the class of complete states such that neuron **A** fires.

Anticipating my treatment of omissions (see section 4), I adopt a similar convention for dealing with events that are essentially a particular neuron's failure to fire:

> For any neuron **A**, such that **A**'s failure to fire at time t is explicitly shown in the neuron diagram, $\neg A$ is the instantaneous event characterised by $(@, t, s, I_{\neg A}, C_{\neg A})$, where
> s = the complete state of @ at t,
> $I_{\neg A}$ = the closed interval $[t, t + dt]$, and
> $C_{\neg A}$ = the class of complete states such that neuron **A** does not fire.

In later chapters, I use these conventions unless otherwise indicated.

3. Recovering features of ordinary events

Instantaneous events differ from ordinary events in two significant ways: first, instantaneous events are global in the sense that they are concerned with complete states, whereas ordinary events are local. Second, instantaneous events are, as their name suggests, instantaneous, whereas ordinary events – such as football matches and birthday parties – are extended over time. In this section, I show how my account of the causal relata can accommodate these features of ordinary events. I begin by showing how there is a derivative sense in which instantaneous events have more specific locations, corresponding to the way we usually assign locations to ordinary events (section 3.1). I then show how my proposed account of the causal relata can accommodate temporally extended events and causal relations between them (section 3.2).

3.1 The location of instantaneous events

We usually think of ordinary events as having quite well-defined spatial locations.[16] Consider, for example, the event of my being out for a walk on the beach: it is natural to think that this event is located where I am – on the beach – and that the region within which it occurs does not include remote regions of space, thousands of light-years away.

On the surface, it might seem that my account of the causal relata cannot accommodate this intuition: on my account, my being out for a walk on the beach at time t corresponds to an instantaneous event *Beach walk* – namely, the instantaneous event characterised by the quintuple $(@, t, s, I, C)$, where s is the complete state of @ at time t, I is some appropriately chosen closed interval, and C is the class of complete states such that I am out for a walk on the beach. The most reasonable and straightforward way to assign a location to this instantaneous event is to say that it encompasses *all* of the actual world at time t – it is *everywhere* where the complete state s is exemplified. In this sense, the instantaneous event of my being out for a walk on the beach is truly global.

There is, however, a derivative sense in which an instantaneous event may be located within a smaller region: as we have seen, instantaneous events are individuated based on what is essential and what is merely accidental to their occurrence. For example, it is essential to the occurrence of *Beach walk* that I am out for a walk on the beach, whereas it is merely accidental that a herd of elephants is splashing in the Okavango Delta. We tend to focus on

[16] Paul and Hall (2013), p. 178.

events based on a class C that is easily specifiable – such as the class C of states such that I am out for a walk on the beach. And when we are concerned with an event based on a particular class C, we tend to associate a unique *specification* with C – such as the specification that I am out for a walk on the beach – even though there are of course many different, more or less disjunctive, ways to specify any given class of complete states.

Whether a state s belongs to a class C is simply a matter of the identity of s. However, whether a state s satisfies the *specification* associated with C is often a matter of what happens within a relatively small region. For example, the fact that a state s is such that I am out for a walk on the beach is grounded in what happens within a relatively small spatial region, namely the small region on the beach where I am in fact walking. This suggests a derivative sense in which instantaneous events have more specific locations: the derivative location of an instantaneous event (@, t, s, I, C) is the region such that the fact that s satisfies the *specification* associated with C is grounded in what happens within this region.[17]

There is nothing metaphysically deep about this notion of derivative locations: this should be clear from the fact that the derivative location of an event depends on which *specification* we associate with the relevant class C. The point of introducing the notion is simply to show how it allows us to recover our ordinary way of thinking about events as having relatively well-defined spatial locations. And indeed, derivative locations correspond nicely to the locations assigned by common sense. In the case of *Beach walk*, for example,

[17] Note that this way of assigning derivative locations is compatible with saying that an event has its derivative location contingently: suppose that the complete state s of @ is such that I am at the northern end of a particular beach at time t. In that case, the fact that s satisfies the relevant specification is grounded in what happens within a small region at the northern end of the beach. However, the event *Beach walk* could have been realized in different ways. For example, it could have been realized by the world being in a complete state s* at the relevant time t such that I am at the southern end of the beach instead. The fact that s* satisfies the specification is grounded in what happens at the southern end of the beach – and a region at the southern end of the beach would therefore be the derivative location of *Beach walk*. In this way, *Beach walk* has its derivative location contingently.

we correctly find that the derivative location of this event is the small region of the beach where I am.

3.2 Temporally extended events, instantaneous events, and piecemeal causation

The second important difference between instantaneous events and ordinary events is this: instantaneous events are, as their name suggests, instantaneous; ordinary events, by contrast, are usually extended over time. Indeed, paradigm examples of ordinary events, such as football matches and birthday parties, go on for hours.

My proposed account of the causal relata needs to be able to accommodate such temporally extended events because our causal statements often relate such extended events. For example, the statement that 'the sunshine melted the snow' – or, more perspicuously, 'the sun's shining caused the melting of the snow' – relates extended events: the extended event of the sun's shining (presumably for several hours), and the extended and temporally overlapping event of the melting of the snow (which, again, presumably extends for several hours). How can my proposed account handle such cases?

My suggested answer comes in two parts. The first part consists in the suggestion that temporally extended events are built up from instantaneous events. For example, the extended event of the sun's shining throughout the morning is built up from a sequence of (presumably infinitely many) instantaneous events – including the instantaneous event of the sun's shining at precisely 10 a.m., the instantaneous event of the sun's shining at precisely 10.01 a.m. etc. And the extended event of the melting of the snow is similarly built up from a sequence of (presumably infinitely many) instantaneous events – including the instantaneous event of snow melting at precisely 10 a.m., the instantaneous event of snow melting at precisely 10.01 a.m. etc.

The second part of my answer consists in the suggestion that causal statements relating extended events are true (when they are true) in virtue of the causal relations between the instantaneous events that make up the relevant extended events. For example, the claim that 'the sun's shining caused the melting of the snow' is true (if it is true) in virtue of the causal relations between instantaneous sunshine events (such as the instantaneous event of the sun's shining at precisely 10 a.m.) and instantaneous melting events (such as the instantaneous event of snow melting at precisely 10.01 a.m.). This suggestion of how the extended event of the sun's shining causes the extended event of the melting of the snow is very similar to what Lewis calls 'piecemeal causation', and following Lewis I will adopt this name.[18]

On Lewis's original analysis of causation, an event *c* causes an event *e* just in case *e* depends causally on *c*, or there is a chain of causal dependence connecting the two (where causal dependence is simply counterfactual dependence between distinct events).[19] Within the context of this analysis, Lewis presents the idea of piecemeal causation as follows:

> 'Suppose that *c* and *e* are large, prolonged processes, each composed of many smaller events. Suppose it is not true (or not clearly true) that *e*, taken as a whole, causally depends on *c*, taken as a whole; suppose even that they are not connected by a chain of causal dependence. It may nevertheless be that *c* and *e* are divisible into parts in such a way that every part of *e* is causally dependent on (or connected by a chain of causal dependence to) some part of *c*. In that case we might well simply speak of *c* as a cause of *e*, though it is not so under the analysis I gave.'[20]

It should be clear that this idea is independent of Lewis's particular analysis of causation: it can be applied in the context of any account of causation. On my proposed conception of the causal relata, what matters is simply this: when we say that an extended event *c* causes an extended event *e*, this claim is true (or

[18] See Lewis (1986*e*), pp. 172-75.
[19] Lewis (1986*d*).
[20] Lewis (1986*e*), p. 172.

false) in virtue of the fact that *c* is made up of instantaneous events and *e* is made up of instantaneous events, such that each instantaneous *c*-event causes some instantaneous *e*-event, and each instantaneous *e*-event is caused by some instantaneous *c*-event.

At first, it might seem cumbersome that my proposed account of the causal relata requires us to understand causal relations between temporally extended events as obtaining in virtue of causal relations between instantaneous events. On closer inspection, however, I believe that this gives a deeper understanding of the causal phenomena. For example, understanding causal relations between extended events in this way makes it entirely straightforward to see how two extended and temporally overlapping events – such as the sun's shining and the melting of the snow – can stand in the relation of cause and effect without any simultaneous or backwards causation.

Furthermore, there are cases where the notion of piecemeal causation is indispensable, even on accounts, such as Lewis's, that do allow temporally extended events to feature among the basic causal relata. Thus, Lewis gives the following example:

> 'Self-sustaining processes exhibit piecemeal causation. For instance, suppose a public address system is turned up until it howls from feedback. The howling, from start to finish, is an event. […] it is not true, on my account, that the howling *taken as a whole* causes itself. What is true is that the howling causes itself piecemeal. It is divisible into parts in such a way that each part except the first is caused by an earlier part, and each part except the last causes a later part. This causing of part by part is unproblematic.'[21]

Once again, thinking about this case in terms of instantaneous events gives a deeper understanding of its causal structure.

[21] Lewis (1986*e*), pp. 172-73.

4. Omissions as causal relata

It seems intuitively clear that omissions can be causes and effects. For example, my thinking about this chapter was a cause of my failure to water my potted primrose, and my failure to water the primrose was in turn a cause of its withering. However, many accounts of the causal relata run into trouble when trying to accommodate omissions.[22] I therefore take it to be an important advantage of my proposal that it can straightforwardly accommodate omissions as being no different in kind from ordinary, positive events, such as my thinking about this chapter or the withering of the primrose.

The trouble with accommodating omissions and absences as causal relata is that it seems that there is nothing to accommodate. As Schaffer writes:

> '[Absences] are *nothings*, non-occurrences, and hence are not in the world.'[23]

My suggestion is that this way of thinking about omissions and absences is a misunderstanding: on my proposed conception of the causal relata, omissions and absences occur in the world in just the same way as ordinary positive events – in fact, there is no deep metaphysical distinction between the two.

To see this, let us begin by considering a simple case – say, my failure at time t to water my potted primrose. Suppose that the actual world is in the complete state s at time t. The complete state s has many different properties – that is, it belongs to many different classes of complete states. For example, it has the property that I'm out for a walk on the beach, that a particular seagull is landing on the railing of a boat, that a herd of elephants is splashing in the Okavango Delta, etc. It also has the property that I am not watering my primrose – that is, it belongs to a class C of complete states such that I do not water my primrose.

[22] See e.g. Beebee (2004); Paul and Hall (2013), pp. 177-83; Schaffer (2016), pp. 8-9.

[23] Schaffer, (2016), p. 8. See also Beebee (2004).

My suggestion is that the instantaneous event of my failing to water the primrose at t has exactly the same realization as every other instantaneous event that actually occurs at t – including the instantaneous event of my being out for a walk on the beach, the instantaneous event of the seagull landing on the railing of a boat, the instantaneous event of the herd of elephants splashing in the Okavango Delta, etc. What distinguishes the instantaneous event of my failing to water the primrose from all these other instantaneous events is its essence: the instantaneous event of my failing to water the primrose at time t consists in the quintuple, $(@, t, s, I, C)$, where I is the closed interval $[t, t+dt]$ with dt being an infinitesimal time-interval, and where C is the class of complete states such that I do not water the primrose.

This instantaneous event is not different from other instantaneous events in any deep metaphysical sense. The only thing that distinguishes it from what we would call positive instantaneous events is the way in which we have picked out the relevant class C of complete states – namely, by reference to what does *not* occur in these states, rather than by reference to what does occur. However, this is merely a superficial difference in our way of picking out the relevant class – C itself is simply a class of complete states, just like any other class of complete states.

So far, we have seen how my account of the causal relata allows us to capture a maximally temporally fragile event of my failure to water the primrose at exactly time t. However, it is natural to understand the statement 'my failure to water the primrose caused it to wither' as a statement relating temporally extended events – namely, the extended event of my failure to water the primrose (e.g. my not watering it throughout a week) and the extended and temporally overlapping event of the primrose withering.

Relying on the idea of piecemeal causation, my suggestion is that the statement that 'my failure to water the primrose caused it to wither' is true in virtue of the causal relations between the instantaneous events that make up these temporally extended events: the temporally extended event of my failure to water the primrose is built up from instantaneous events of my failure to water it – namely, instantaneous events of omission of the kind discussed above. Similarly, the temporally extended event of the primrose withering is built up from instantaneous events of flowers and leaves losing turgor pressure, wilting and subsequently drying out. And the statement is true in virtue of the fact that each instantaneous event of my failure to water the primrose – together with the already dry soil in the pot and progressively weaker state of the primrose – causes subsequent instantaneous withering events, and each instantaneous withering event is caused by an instantaneous event of my failure to water it.[24]

To bring out the advantages of this way of accommodating omissions, we may contrast it with a different view that is sometimes suggested: that cases such as the above can be handled by taking omissions to be 'ordinary events, oddly described'.[25] On this view one might say that my failure to water my primrose at *t* just is my being out for a walk at *t*. Lewis considers this option in the following passage:

> 'Fred omits the precautions, sleeping through the time when he was supposed to attend to them. His nap was a genuine event; it is not objectionably disjunctive. There are many and varied ways in which he could have omitted the precautions, but there is just one way that he did omit them. We could plausibly say, then, that his nap *was* his omission of precautions.'[26]

There are a number of problems with this view – for example, it comes into conflict with the independently plausible principle that causation is an extensional relation (for more on extensionality, see section 5).[27] Furthermore, there are cases in which it is simply not clear which ordinary event the

[24] Note that the example is fictive. No plants were harmed in the writing of this dissertation.
[25] Paul and Hall (2013), p. 180; cf. Lewis (1986*e*), p. 192-3. Note that Paul and Hall do not endorse this view.
[26] Lewis (1986*e*), p. 192.
[27] Paul and Hall (2013), p. 180.

omission or absence should be identified with. For example, Paul and Hall present the following case:

> 'suppose that overpopulation of rabbits in a certain region, at a certain time, is caused in part by a lack of predators. Are we really to suppose that there are some (potential) predators in that region at that time, doing something else other than preying on these rabbits – and that it is that something else we are referring to by the expression "a lack of predators"? At best, we might claim that the region of the rabbits, just as it is, counts as the event described by "the lack of predators," but this again is unsatisfactory.'[28]

Cases such as the above present a severe challenge for the view that omissions and absences should simply be understood as ordinary events oddly described.[29]

By contrast, my proposed account of the causal relata can straightforwardly accommodate the above case: suppose that the world is in the complete state s at time t. On my account, then, the instantaneous event consisting in the absence of predators at time t has exactly the same realization as every other instantaneous event occurring in @ at t – namely, the realization given by the triple (@, s, t). What distinguishes the instantaneous event of the lack of predators from all the other instantaneous events occurring at the same time is its essence: the instantaneous event of the lack of predators at time t corresponds to the quintuple (@, t, s, I, C), where I is the closed interval $[t, t+dt]$ and C is the class of complete states such that there are no predators in the relevant region.

---

[28] Paul and Hall (2013), p. 181.

[29] Schaffer's suggestion that causation is a quaternary, contrastive relation (Schaffer (2005)) may be seen as a sophisticated version of this view. On Schaffer's suggestion, a description such as 'my failure to water the primrose' is an odd way of describing what I was in fact doing at the time, i.e. going for a walk on the beach. This odd description draws attention to a particular contrast to my walking on the beach – namely, my watering the primrose. It is not entirely clear how Schaffer's proposal can handle more challenging cases such as the lack-of-predators case quoted above.

Again, what we have characterised here is the maximally temporally fragile event of the absence of predators at exactly time *t*. When we say that the lack of predators caused the overpopulation of rabbits, however, it is plausible that we are talking about a temporally extended event of there being a lack of predators throughout several months or more (after all, an instantaneous absence would not have much of an effect on its own). By contrast, we may if we like take the overpopulation of rabbits to be an instantaneous event, consisting simply in there being too many rabbits in the area at a given time. On this interpretation, the statement that 'the absence of predators caused the overpopulation of rabbits' gets to be true (if it is true) in virtue of the relations that hold between instantaneous absence-of-predators events and the instantaneous overpopulation event: each instantaneous absence-of-predators event is – together with the presence of parent rabbits, litters of baby rabbits, ample food, etc. – a cause of a subsequent instantaneous event of parents and litters of baby rabbits surviving yet another instant of carefree nibbling. And through a long chain this eventually causes the instantaneous event of the overpopulation of rabbits.

As a final example of how my account of the causal relata can accommodate omissions and absences, let us consider a case that seems to present even greater difficulties – namely, the case of a void. Lewis gives the following colourful example:

> 'The void is deadly. If you were cast into a void, it would cause you to die in just a few minutes. It would suck the air from your lungs. It would boil your blood. It would drain the warmth from your body. And it would inflate enclosures in your body until they burst.'[30]

It seems intuitively true, as Lewis says above, that a void can cause things. However, a void is nothing at all:

---

[30] Lewis (2004*b*), p. 277.

> 'if there is a void within these walls, then (even though the walls are some distance apart) there is nothing at all between the walls. What? – Not even any spacetime? Not even any flat, causally inert spacetime? – No, not even any spacetime. Nothing at all.'[31]

In this way, the above case presents trouble even for a view that identifies omissions and absences with spatiotemporal regions. For in the above case, there *is* no spatiotemporal region with which we may identify the void. By contrast, my proposed account of the causal relata can accommodate the above case just as straightforwardly as the others we have considered: treating the world of the example as actual, the first relatum when we say that the void causes the test person's death is an instantaneous event of our usual kind – namely, the quintuple ($@$, t, s, I, C), where s is the complete state of $@$ at the relevant time t, I is the closed interval $[t, t+dt]$, and C is the class of complete states such that there is a void within the relevant boundaries.

Indeed, as far as I can see, the only kind of case that I cannot accommodate is a case in which there is no world at the relevant time t. In that case, there can be no instantaneous events occurring at t: when there is no world, there is indeed *nothing* – and there is nothing that can play the role of causal relatum. But this is of course just the result we want: causation is something that happens within a world – so if there is no world, there can be no causation.

5. Causal differences and extensionality

A second challenge for accounts of the causal relata arises from the observation that our causal judgements seem to be sensitive to very fine differences. To take an example from Lewis, suppose that John says "Hello",

[31] Lewis (2004*b*), p. 278.

and he says it 'rather too loudly'.[32] In this case, it seems that John's saying "Hello", and John's saying "Hello" loudly must be different causal relata, since they have different causes and effects: John's state of tension is a cause of his saying "Hello" loudly, but not of his saying "Hello" as such; and John's saying "Hello" is a cause of Fred's greeting him in return, while John's saying "Hello" loudly is not.[33]

On my proposed account of the causal relata, we can easily accommodate cases such as this: treating the world of the example as actual, my suggestion is that the instantaneous event of John's saying "Hello" and the instantaneous event of John's saying "Hello" loudly have the same realization – characterised by the triple $(@, s, t)$, where s is the complete state of $@$ at the relevant time t. However, the two events have different essences: the instantaneous event of John's saying "Hello" is essentially a saying "Hello", but only accidentally a saying "Hello" loudly; by contrast, the instantaneous event of John's saying "Hello" loudly is essentially a saying "Hello" loudly. This solution – taking the difference between the two events to be a difference in essence – is exactly what Lewis himself proposes:

> 'Arguably there is one event that occurs which is essentially a saying -"Hello" and only accidentally loud; it would have occurred even if John had spoken softly. Arguably, there is a second event that implies, but is not implied by, the first. This event is essentially a saying -"Hello"- loudly, and it would not have occurred if John had said "Hello" but said it softly.'[34]

The desideratum that an account of the causal relata should be able to handle cases involving causal differences such as the above is closely related to the desideratum that an account should allow us to maintain that causation is an extensional relation:

---

[32] Lewis (1986*g*), p. 255.
[33] Lewis (1986*g*), p. 255.
[34] Lewis (1986*g*), p. 255.

To say that causation is an extensional relation is to say that if *c* causes *e* and $c = c^*$ and $e = e^*$, then c^* causes e^*. At the semantic level, this means that different descriptions can be substituted for each other without changing the truth-value of our causal claims, provided the descriptions refer to the same event.[35] It seems intuitively plausible to hold that causation is an extensional relation. As Strawson writes:

> 'causality is a natural relation … that relationship holds however *A* and *B* may be described.'[36]

However, maintaining that causation is an extensional relation imposes severe constraints on accounts of the causal relata. This is brought out by examples such as the following:

> 'one might accept that McEnroe's tension caused his serving awkwardly but deny that his tension caused his serving. One wants to say: he was going to serve anyway.'[37]

If we were to say that McEnroe's serving is the same event as his serving awkwardly, the above case would present a counterexample to the extensionality of causation: 'McEnroe's serving' cannot be substituted for 'McEnroe's serving awkwardly' while preserving the truth-value of the claim that 'his tension caused McEnroe's serving awkwardly'. To maintain that causation is an extensional relation – while respecting our intuitive causal judgements – we therefore need an account where the causal relata are individuated sufficiently finely to ensure that e.g. McEnroe's serving is numerically distinct from McEnroe's serving awkwardly.

---

[35] Schaffer (2005), p. 306.
[36] Strawson (1985), p. 118.
[37] Schaffer (2005), p. 307.

My account of the causal relata easily achieves this: on my account, the two instantaneous events of McEnroe's serving and McEnroe's serving awkwardly have the same realization – namely, the realization given by the triple (@, *t*, *s*), where *s* is the complete state of @ at the relevant time *t*. However, the two events differ in their modal profiles: the instantaneous event of McEnroe's serving is essentially a serving, but only accidentally awkward; McEnroe's serving awkwardly is a more fragile version of this event – it is essentially McEnroe's serving awkwardly.

In addition to cases such as the above, there is another group of cases that seem to present even more challenging counterexamples to the extensionality of causation – namely, cases that involve mere focal differences. For example, it seems correct to say that Socrates's *drinking hemlock* at dusk caused his death, whereas it seems false to say that Socrates's drinking hemlock *at dusk* caused his death.[38] As Schaffer notes: 'One wants to say: *when* he drank the hemlock did not matter.'[39]

My treatment of such cases has two parts. First, we need to note that we may use both essential and accidental features of events to pick them out. Thus, the precise way in which we describe an instantaneous event gives hints, but is *not* an infallible guide to what is essential and what is merely accidental to the event we want to be talking about.[40] In particular, describing an instantaneous event as 'Socrates' drinking hemlock at dusk' does not by itself tell us whether Socrates' drinking hemlock, or occurring at dusk, is essential to the event we want to be talking about.

This, I suggest, is where mere focal differences can in fact matter in specifying which instantaneous event we are referring to: the emphasis on 'drinking hemlock' in 'Socrates' *drinking hemlock* at dusk' suggests that Socrates's drinking hemlock is essential, whereas the time of occurrence may be merely

---

[38] Schaffer (2005), p. 307. The example is originally due to Achinstein (1975).
[39] Schaffer (2005), p. 307.
[40] See e.g. Lewis (1986*g*), pp. 247-54; Paul and Hall (2013), pp. 102-3; Yablo (1992*b*), note 28, pp. 439-40.

accidental. By contrast, the emphasis on 'at dusk' in 'Socrates' drinking hemlock *at dusk*' suggests that the time of occurrence is essential. If this is correct, then we can see how mere focal differences can indeed make a difference as to which instantaneous event we are referring to.

To fully capture our causal judgements in this case, however, it is plausible that we need to adopt Schaffer's suggestion that focal differences should be understood as contrastive differences:

> 'the focus effect begs for contrastive explanation. Thus 'Socrates' *drinking hemlock* at dusk' is naturally interpreted as *c*: Socrates' drinking hemlock at dusk, rather than *c**: Socrates' drinking wine at dusk (or some other salient alternative to drinking hemlock); whereas 'Socrates' drinking hemlock *at dusk*' is naturally interpreted as *c*: Socrates' drinking hemlock at dusk, rather than *c**: Socrates' drinking hemlock at dawn (or some other salient alternative to occurring at dusk).'[41]

In Chapter 10 section 3, I show how my proposed account of causation can accommodate such contrastive causal statements.

6. Transitivity

As a third and final application of my proposed account of the causal relata, I will now show how it gives us the resources to withstand a particular kind of counterexample to transitivity.

At first glance, one might wonder why we should care about this application: as we have already seen in Chapter 2 section 1.3.2, there are convincing counterexamples to the transitivity of causation. It might therefore seem natural to suggest that since, as Schaffer writes, 'transitivity is lost anyway',[42] it simply does not matter whether our account of the causal relata allows us to withstand counterexamples to transitivity of this particular kind.

[41] Schaffer (2005), p. 308.
[42] Schaffer (2016), p. 14.

However, this suggestion is too hasty: transitivity plays an important role in our causal reasoning. Even though we cannot maintain an unrestricted principle of transitivity, it is therefore an important desideratum that we should be able to maintain a *restricted* transitivity principle (see Chapter 11 section 2). And to achieve that goal, we should aim for an account of the causal relata that limits the range of counterexamples to transitivity as much as possible.

With this motivation, let us now consider the following example of the kind of case that our account of the causal relata should enable us to handle:

> *Skiing accident*: while skiing, Suzy breaks her right wrist. The next day, she writes a philosophy paper, which is subsequently accepted for publication. Since Suzy's right wrist is broken, she writes the paper by typing with her left hand. And as she is not used to writing this way, she develops a cramp in her left hand.[43]

On an account where the causal relata are not sufficiently finely individuated, this case presents a counterexample to transitivity. To see this, consider an account on which the event of Suzy's writing the paper is identical to the event of Suzy's writing the paper by typing with her left hand. Now, it is clear that Suzy's skiing accident causes Suzy's writing the paper by typing with her left hand. From the assumption that the event of Suzy's writing the paper is identical to the event of Suzy's writing the paper with her left hand, it follows by substitution that Suzy's skiing accident causes Suzy's writing the paper. This is already a bad result, showing how identifying Suzy's writing the paper with Suzy's writing the paper with her left hand brings us into conflict with the principle that causation is an extensional relation. And from here, there is only a small step to the full counterexample to transitivity: it is clear that Suzy's writing the paper is a cause of the paper's being accepted for publication. By

[43] This case is closely based on a case presented in Paul (2004*a*). For discussion of similar cases, see also Ehring (2009), pp. 403-4; McDonnell (forthcoming); Paul and Hall (2013), pp. 237-44; Schaffer (2016), pp. 13-14, and Woodward (1984), pp. 234-46.

transitivity, we therefore find that Suzy's skiing accident causes the paper's getting accepted for publication. But that seems false – intuitively, Suzy's skiing accident does *not* cause the paper's being accepted for publication.

One might try to resist this result by denying that Suzy's skiing accident causes Suzy's writing the paper by typing with her left hand. However, that option is untenable. For it seems intuitively obvious that Suzy's skiing accident is a cause of the cramp in Suzy's left hand – indeed, if the accident had not happened, Suzy would not have had the cramp. And of course, the skiing accident does not cause the cramp through some spooky action at a distance – it does so with Suzy's writing with her left hand as an intermediary.[44] As long as the event of Suzy's writing the paper is assumed to be identical to the event of Suzy's writing the paper with her left hand, the counterexample to transitivity therefore remains.

My proposed account of the causal relata easily provides the resources needed to handle this purported counterexample to transitivity, since it individuates events based on their modal profile. On my account, we therefore find that there are two instantaneous writing events occurring at the relevant time – a writing that is essentially a writing of the paper, but only accidentally a writing by typing with the left hand; and a more fragile version of this event that is essentially a writing of the paper by typing with the left hand. These instantaneous events have the same realization, but different modal profiles – and they are therefore numerically different events.

Since my account of the causal relata yields the result that Suzy's writing the paper is *not* identical to Suzy's writing the paper by typing with her left hand, we now have the resources to resist the counterexample. Here, in brief, is what we should say: Suzy's skiing accident is *not* a cause of Suzy's writing her paper, though it is a cause of the more fragile event of Suzy's writing her paper by typing with her left hand. In turn, the robust event of Suzy's writing her paper is a cause of the paper's getting accepted for publication, while the more

[44] Paul (2004*a*), pp. 209-10; cf. Paul and Hall (2013), p. 238.

fragile event of Suzy's writing her paper by typing with her left hand is not. Thus, transitivity simply does not apply. Indeed, the only legitimate application of transitivity in the present case is the following: the more fragile event of Suzy's writing her paper by typing with her left hand is a cause of Suzy's cramp. By transitivity, it follows that Suzy's skiing accident is a cause of Suzy's cramp. That is exactly the result we want. In this way, my proposed account of the causal relata gives us the resources we need to resist the counterexample.

It should be clear, however, that what has been shown here is relatively limited: all I have shown is that my account of the causal relata gives us the *resources* we need to resist the counterexample, by allowing us to say that the event of Suzy's writing the paper is *not* identical to the event of Suzy's writing the paper by typing with her left hand. To truly handle the case, however, we also need an account of causation that delivers the intuitively correct verdicts – e.g. that Suzy's skiing accident is *not* a cause of Suzy's writing the paper, though it is a cause of the more fragile event of Suzy's writing the paper by typing with her left hand. In Chapter 7 section 4.3, I therefore return to the case and show that my proposed account of causation does indeed deliver these verdicts. For now, however, we have established what we need – namely, that my account of the causal relata provides the needed resources.

7. Restricting the domain of quantification

It follows from the principle of *Plenitude* that there is a great multitude of instantaneous events. Common sense accepts some of these events, but rejects others.

Suppose, for example, that Fred is out for a walk on the beach at time *t*. By *Plenitude*, there is a multitude of instantaneous events that all occur at time *t* and have the same realization – namely, the realization given by the complete state of the world at *t* – but which differ in their modal profiles. Among these is the instantaneous event that is essentially Fred's walking. But in addition, there are also instantaneous events based on more gerrymandered properties of the actual complete state of the world – such as the event that is essentially Fred's walking *or* a bumblebee landing on a clover blossom. Common sense recognises some of these events, but rejects others. For example, it recognises the event that is essentially Fred's walking. But it rejects events based on more gerrymandered properties, such as the event that is essentially Fred's walking *or* a bumblebee landing on a clover blossom.

When we quantify over events, we typically restrict our domain of quantification so as to exclude events with overly disjunctive essences, such as the event that is essentially Fred's walking *or* a bumblebee landing on a clover blossom. However, it may well depend on context exactly what counts as 'overly disjunctive'. In some contexts, for example, we may accept the event that is essentially someone's watering my flowers; in other contexts, we may reject this as having an overly disjunctive essence, and accept only the less disjunctive events of Alex's watering my flowers, Fenner's watering my flowers, etc.

It is important to recognise that we restrict our domain of quantification in this way: my proposed account of causation requires us to quantify over instantaneous events. To apply the account, we therefore need to specify the relevant domain of instantaneous events. Indeed, a natural way to capture this is to understand the contextually determined domain of instantaneous events as a further causal relatum. In the following, however, I will leave this relativisation to a domain of instantaneous events implicit: as it turns out, the contextually relevant domain of instantaneous events remains remarkably stable across a wide range of different contexts. Once we have an understanding of the relevant domain, we can therefore let the choice of domain recede into the background while we attend to other features of the causal relation.

The following is a brief characterisation of our typical choices of domain of quantification, which I will presuppose in the following:

When we are dealing with neuron diagrams, the domain of quantification contains – except in very special contexts – only instantaneous events that are maximally temporally fragile, and based on the behaviour of a single neuron. Thus, our domain of quantification when dealing with the case of early preemption in Figure 1 contains just the five events A, B, C, $\neg D$, and E, defined in accordance with the conventions set out in section 2.3 above.

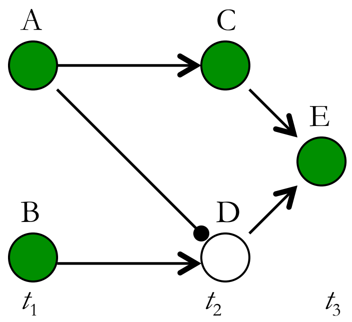


Figure 1

When we are dealing with real life cases, we can expect to find a bit more contextual variation in the domain of quantification. In general, however, the following seems to hold: we may think of any class C of complete states as inducing a partition on all complete states – namely, the partition into C and the complement of C. Let us say that a class C is *good*, just in case it induces a reasonably joint-carving partition – i.e. just in case either C or the complement of C captures a class of complete states that are genuinely similar to each other. Then my suggestion is that we include an event e within our domain of quantification just in case e is based on a *good* class C.

This suggestion leaves room for contextual variation: it may well depend on context what the relevant standard of goodness is. At the same time, the suggestion captures how the standards we apply to omissions and absences seem to be a mirror image of the standards we apply to so-called positive events: since what matters is the *partition* induced by a class C, we find that a class C is good just in case the complement of C is also good. This correctly captures the fact that in any context where we include the event of my watering the primrose within our domain of quantification, we also include the event of my failure to water the primrose, and *vice versa*.

Paying attention to the partition induced by a particular class *C* thus allows us to see that there is a principled basis for including omissions and absences within our domain of quantification, while excluding events such as the event that is essentially Fred's walking *or* a bumblebee landing on a clover blossom:

It is indeed true that omissions and absences have highly disjunctive essences. As Lewis writes:

> 'an event of omission, essentially specifiable as such, is highly disjunctive. Fred omits the precautions if he does something else during the period in which he was supposed to attend to them. So there are as many different ways for the event of omission to occur as there are alternative ways for Fred to spend the time. An event essentially specifiable as an omission amounts to an event essentially specifiable as a sleeping-or-loafing-or-chatting-or- … with a disjunct for everything Fred might do other than attending to the precautions.'[45]

However, when a class *C* characterises an omission or absence, then the *complement* of *C* characterises a class of states that are genuinely similar. For example, the class characterising Fred's failure to attend to the precautions is the class of states such that Fred does attend to the precautions – and the complete states in this class are genuinely similar. In this way, the class of states characterising an omission or absence induces a sensible *partition* on all the complete states. By contrast, the class *C* characterising the objectionably disjunctive event of Fred's walking *or* a bumblebee landing on a clover blossom does not induce a sensible partition: neither *C* itself nor the complement of *C* captures a class of complete states that are genuinely similar.

[45] Lewis (1986), 'Causation', p. 190.

8. Conclusion

In this chapter, I have presented an account of the causal relata on which the causal relata are *actual instantaneous events*, characterised by a realization (@, t, s) and a modal profile (I, C). Here @ is the actual world, t is a time, s is the complete state of @ at time t, I is a closed interval that includes time t, and C is a class of complete states that includes s.

This account has two important advantages. First, it harmonises our ordinary understanding of the causal relata with the conception of the laws of nature that one finds in the philosophy of science: instantaneous events are built from the very same global and instantaneous complete states whose forward evolution is governed by the laws of nature. At the same time, my treatment of derivative locations and temporally extended events allows us to recover our ordinary understanding of the causal relata as being local and temporally extended.

Second, my proposed account gives us the resources to handle the standard test cases for accounts of the causal relata: it allows us to accommodate omissions and absences as causal relata, it allows us to respect causal differences and preserve extensionality, and it allows us to resist a particular kind of counterexample to the transitivity of causation.

The real test of an account of the causal relata, however, is whether it can underlie a successful account of causation.[46] Apart from the considerations mentioned above, one of my main motivations in developing the account presented here has been to develop an account of the primary causal relata tailored specifically to my proposed account of causation. Before I get to the account of causation itself, however, I first need to introduce a further kind of causal relatum – namely, possibility horizons.

[46] Cf. Stein (2016), p. 482.

5

The third relatum: possibility horizons

My suggestion is that causation is a ternary relation: *c* causes *e* within possibility horizon $\mathscr{H}$. The notion of a possibility horizon is inspired by the following quotation from Mackie:

> '[Causal] statements are commonly made in some context, against a background which includes the assumption of some *causal field*. A causal statement will be the answer to a causal question, and the question 'What caused this explosion?' can be expanded into 'What made the difference between those times, or those cases, within a certain range, in which no such explosion occurred, and this case in which an explosion did occur?' Both causes and effects are seen as differences within a field; anything that is part of the assumed (but commonly unstated) description of the field itself will, then, be automatically ruled out as a candidate for the role of cause.'[1]

In brief, a possibility horizon is a restricted class of nomologically possible worlds – namely, a class consisting in the actual world together with those nomologically possible worlds that represent *relevant possibilities*. My aim in this chapter is to develop this notion of a possibility horizon in more detail.

By suggesting that causation is a ternary relation, with a possibility horizon as its third relatum, I am taking a stand on a rather contentious issue: namely, how we should account for the context-dependence of causal claims. It is widely recognised that the assertability of causal claims depends on

[1] Mackie (1974), pp. 34-35.

context.[2] However, it is debated whether this context-dependence should be given a semantic or a merely pragmatic explanation. For a long time, the orthodox view has been that the context-dependence of causal claims is a rather superficial phenomenon that should be given a merely pragmatic explanation. This view is expressed, for example, in the following quotation from Lewis:

> 'We sometimes single out one among all the causes of some event and call it "the" cause, as if there were no others. Or we single out a few as the "causes", calling the rest mere "causal factors" or "causal conditions." […] We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking).'[3]

Against this, Schaffer and others have suggested that the context-dependence of causal claims goes deeper, and requires a semantic explanation: merely pragmatic considerations cannot account for the ways in which the assertability of causal claims depends on context.[4]

My suggestion that causation is a ternary relation, with a possibility horizon as its third relatum, sides with Schaffer on the question of how to accommodate the context-sensitivity of causal claims: it accommodates the context-sensitivity of causal claims by building context-sensitivity into the semantics. In brief, my suggestion works as follows: metaphysically, causation is a ternary relation. For any two instantaneous events *c* and *e*, and any

[2] See e.g. Maslen (2004), and Schaffer (2012).

[3] Lewis (1986*d*), p. 162. Cf. Lewis (1986*f*), pp. 215-16.

[4] Schaffer (2012), pp. 42-44; see also Maslen (2004); Northcott (2008); Schaffer (2005).

possibility horizon $\mathscr{H}$, there is a context-independent fact of the matter as to whether *c* causes *e* within $\mathscr{H}$.[5]

Usually, however, the surface form of our causal claims is binary: the standard form of a causal claim is simply '*c* causes *e*'. The relevant possibility horizon is then supplied by context – and this explains the context-sensitivity of causal claims: in one context, the causal claim '*c* causes *e*' may be completed as '*c* causes *e* within $\mathscr{H}_1$'; in a different context, it may be completed as '*c* causes *e* within $\mathscr{H}_2$'. And these two complete causal claims may obviously have different truth-values.

In this chapter, I develop the notion of a possibility horizon in more detail. I begin by setting out a definition of what a possibility horizon is (section 1). From a metaphysical perspective, this is all we need. To evaluate ordinary causal claims of the form '*c* causes *e*', however, we also need to know how the context of a causal inquiry selects a possibility horizon. My answer to this question builds on the distinction between default and deviant states. I propose a way to give a reductive characterisation of the default-deviant distinction (section 2). And based on this, I set out a recipe for how context selects a possibility horizon (section 3). Finally, I show how such a contextually determined possibility horizon perfectly models the intuitive distinction between causes and background conditions (section 4).

[5] As mentioned in Chapter 4 section 7, context also imposes restrictions on the domain of instantaneous events. More precisely, we should therefore say that for any two instantaneous events *c* and *e*, possibility horizon $\mathscr{H}$, and domain of instantaneous events, there is a context-independent fact of the matter as to whether *c* causes *e* within $\mathscr{H}$, given the relevant domain of instantaneous events. This more precise statement makes it explicit that the contextually determined domain of instantaneous events functions as a fourth causal relatum. In the following, however, I leave the relativisation to a contextually determined domain implicit, since there seems to be agreement about the relevant domain of instantaneous events across a wide range of contexts.

1. Defining possibility horizons

Formally, a possibility horizon is simply a class of nomologically possible worlds that includes the actual world @.

I follow the standard characterisation of nomologically possible worlds, where a world is nomologically possible just in case it is governed by the actual laws of nature (where the world of the example under consideration is treated as actual). Furthermore, I allow that a nomologically possible world can start out in any complete state – what is required is simply that it evolves forward from then in accordance with the actual laws of nature.[6] Based on this, we may define a possibility horizon as follows:

> *Possibility horizon*: a set of worlds W characterises a possibility horizon iff
> a) W includes the actual world @, and
> b) W includes only nomologically possible worlds.

A possibility horizon $\mathscr{H}$ characterises a restricted modality: a proposition *p* is possible within $\mathscr{H}$ just in case *p* is true in at least one world in $\mathscr{H}$, and a proposition *p* is necessary within $\mathscr{H}$ just in case *p* is true in all worlds in $\mathscr{H}$.

We may think of a possibility horizon $\mathscr{H}$ as representing differences within a causal field of the kind that Mackie describes in the quote above: the worlds included in $\mathscr{H}$ are just those worlds that represent *relevant possibilities* in the given context. Thus, a proposition *p* is possible within $\mathscr{H}$ just in case it expresses a relevant possibility, and a proposition *p* is necessary within $\mathscr{H}$ just in case its truth can be taken for granted in the given context – that is, just in case its negation does *not* express a relevant possibility.

This fits nicely with the formal characterisation of a possibility horizon given above: a possibility horizon must always include the actual world @, since @ always represents a relevant possibility. And a possibility horizon must never include nomologically impossible worlds, since – for the purposes of

[6] Cf. Paul and Hall (2013), pp. 72-73.

determining what causes what – these worlds never represent relevant possibilities.

To make the notion of a possibility horizon more concrete, consider the following case of causation by omission:

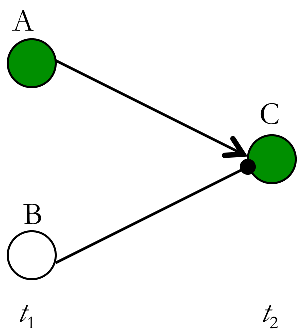


Figure 1

Let us treat the world of the example as actual, and denote it by '@'. From the standpoint of @, there are four nomologically possible worlds, corresponding to the different combinations of whether **A** and **B** fire or fail to fire at time t_1.[7] These four nomologically possible worlds are:

	Complete state at time t_1:
@:	**A** fires, **B** does not fire
w_1:	**A** does not fire, **B** does not fire
w_2:	**A** fires, **B** fires
w_3:	**A** does not fire, **B** fires

For each of these four worlds, it follows from the principle of *Modal profile* (see Chapter 4 section 2.1) which of the three events A, $\neg B$, and C occur (where A is essentially **A**'s firing at t_1, $\neg B$ is essentially **B**'s failure to fire at t_1, and C is

[7] I here adopt the simplified treatment of the case that I outlined in Chapter 3 section 4, where I proceed as if the complete state of @ at time t_1 can be specified simply by specifying the behaviour of **A** and **B**. Furthermore, we only need to consider what happens in the worlds under consideration at time t_1 and t_2. Thus, we may safely limit ourselves to considering only worlds that start out at time t_1 – what happens earlier is irrelevant. I adopt this simplification in all the following cases.

essentially **C**'s firing at t_2, in accordance with the conventions set out in Chapter 4 section 2.3). For convenience, I will in the following include a right-hand column indicating the occurrence or non-occurrence of salient events, where events that do *not* occur are in grey ink:

	Complete state at time t_1:	*Events at t_1*:	t_2:
@:	**A** fires, **B** does not fire	A, $\neg B$	C
w_1:	**A** does not fire, **B** does not fire	A, $\neg B$	C
w_2:	**A** fires, **B** fires	A, $\neg B$	C
w_3:	**A** does not fire, **B** fires	A, $\neg B$	C

The class of worlds W = {@, w_1, w_2, w_3} characterises a possibility horizon according to my above definition, and so does every subset of W that includes @. Thus, {@}, {@, w_1}, {@, w_2}, etc., all characterise possibility horizons.

2. A non-causal characterisation of the default-deviant distinction

How does context select a possibility horizon? My suggestion, which I develop in section 3, is that it does so on the basis of the default-deviant distinction. My aim in this section therefore is to give a non-causal characterisation of the distinction. I begin by presenting the distinction as it is standardly characterised (section 2.1). I then suggest an alternative way of capturing the distinction that does not rely on causal notions (section 2.2).

2.1 The default-deviant distinction as standardly characterised

It has often been noted that the distinction between default and deviant states – or some distinction along those lines – seems to play an important role in our thinking about causation.[8] Hall characterises the default-deviant distinction as follows:

[8] See e.g. Hall (2007*a*) and (2007*b*); Hitchcock (2007); Halpern and Hitchcock (2015); and, for related proposals, Maudlin (2004) and Menzies (2004). For critical discussion, see Blanchard and Schaffer (forthcoming).

> 'we very often find, in contemplating various parts of the world, that we have a reasonably clear and firm conception of what that part would be doing if nothing was happening to it. That is its default state; anything else counts as a deviation.'[9]

Hitchcock gives a very similar characterisation:

> 'As the name suggests, the default value of a variable is the one that we would expect in the absence of any information about intervening causes. More specifically, there are certain states of a system that are self-sustaining, that will persist in the absence of any causes other than the presence of the state itself: the default assumption is that a system, once it is in such a state, will persist in such a state.'[10]

In addition, Hitchcock characterises the default-deviant distinction through the following rules of thumb:

> 'Temporary actions or events tend to be regarded as deviant outcomes. In the case of human actions, we tend to think of those states requiring voluntary bodily motion as deviants and those compatible with lack of motion as defaults. In addition, we typically feel that deviant outcomes are in need of explanation, whereas default outcomes are not necessarily in need of explanation. Frequently, but not always, my deviant values correspond to positive events, and defaults correspond to absences or omissions.'[11]

This, I believe, gives an initial feel for the default-deviant distinction. There are, however, two problems that need to be resolved before we can use the distinction to characterise how context selects a possibility horizon.

---

[9] Hall (2007*b*), p. 46.
[10] Hitchcock (2007), p. 506.
[11] Hitchcock (2007), p. 507.

The first problem is relatively minor: Hall and Hitchcock characterise the default-deviant distinction as it applies, respectively, to *states* and *values of variables* within a causal modelling framework. To apply the distinction to possibility horizons, I need to extend it to *actual instantaneous events*. However, this is easily done as follows: an actual instantaneous event *e* is a default event just in case we would expect it to occur in the absence of any information about intervening causes.

The second problem is that the standard characterisations of the default-deviant distinction rely on explicitly causal notions.[12] On Hall's definition, the default state is characterised as what the relevant part of the world 'would be doing *if nothing was happening to it*'.[13] On Hitchcock's definition, the default value of a variable is 'the one that we would expect in the absence of any information about intervening *causes*'.[14]

Strictly speaking, my account of how context selects a possibility horizon is not part of my account of what causation is out in the world: my suggestion is that causation out in the world is a ternary relation between two instantaneous events and a possibility horizon. Relying on a causal characterisation of how context selects a possibility horizon is therefore compatible with may aim of giving an ontological reduction of the ternary relation of causation.[15]

As we shall see, however, a non-causal characterisation of the default-deviant distinction – and, by extension, of the way in which context selects a possibility horizon – is in fact available. This characterisation of the default-deviant distinction is revealing in itself. And furthermore, it allows me to give

---

[12] See e.g. Hall (2007*b*), p. 47; Paul and Hall (2013), pp. 51-53 and 200-211.

[13] Hall (2007*b*), p. 46. My italics.

[14] Hitchcock (2007), p. 506. My italics.

[15] It is, for example, a perfectly acceptable suggestion that when someone makes a binary causal claim of the form '*c* causes *e*', a principle of charity is at work, and context therefore – if possible – selects a possibility horizon $\mathscr{H}$ such that the completed causal claim '*c* causes *e* within $\mathscr{H}$' comes out true. I am grateful to Derek Ball for making me aware of this point.

an account of how context selects a possibility horizon that avoids any suspicion that we might be smuggling in causal notions where they do not belong.

2.2 Extrapolation and the default-deviant distinction

My suggestion is that the key to giving a non-causal characterisation of the default-deviant distinction is to be found in the notion of *extrapolation*.[16]

We may think of extrapolation as a kind of induction: when we extrapolate, we recognise a pattern and extend this pattern to cover as yet unobserved times, places, etc. Importantly, extrapolation, understood in this way, is a *non-causal* notion: it is simply concerned with the recognition and extension of patterns. To reinforce this point, note that extrapolation in this sense can be applied in cases where there simply *is* no underlying causal structure: for example, we may use extrapolation to extend a mathematical graph, understood as a purely abstract, non-causal pattern.

Even so, one might well object that bringing in the notion of extrapolation – with its close connection to induction – takes us from the frying pan into the fire: it is well known that induction is connected with deep and intractable problems. In particular, it is connected with Hume's problem of induction, which amounts to the question of how we can give a non-circular justification of our practices of induction.[17] And in addition, it is connected with Goodman's so-called new problem of induction, which amounts to the question of why certain adjectives – such as 'green' – form a suitable basis for induction, while others – such as 'grue' – do not (an object is 'grue' just in case it is examined before a certain time *t* and is green, or it is *not* examined before time *t* and is blue).[18]

However, my proposal for understanding the default-deviant distinction

[16] I here use 'extrapolation' in a broad sense, where interpolation is taken to be a subspecies of extrapolation.

[17] Hume (2007), pp. 61- 65. (Book I Part III section VI).

[18] Goodman (1954), especially pp. 73-75.

in terms of extrapolation does not require us to resolve these problems. Instead, all we need is the *descriptive* fact that we *do* have well-established practices of induction and, in particular, extrapolation. This descriptive fact should be relatively uncontroversial: for example, there is little doubt that, as a matter of descriptive fact, we recognise 'green' as a suitable basis for induction, and reject 'grue'.

Relying on such purely descriptive facts about our practices of induction and, in particular, extrapolation, we may now think of extrapolation as something that yields predictions about which events will occur in a particular region from a particular time *t* onwards. On this basis, we may think of the distinctive feature of a default event as being that all our extrapolations are compatible with the occurrence of this event. By contrast, the distinctive feature of a deviant event is that some of our extrapolations yield predictions that are incompatible with its occurrence.

Suppose, for example, that up until just before time *t*, Suzy is standing and chatting with Billy (without throwing any rocks). In the absence of any information about what happens next, extrapolation yields the prediction that Suzy keeps standing there, chatting with Billy (without throwing any rocks). If Suzy is in fact standing there, chatting with Billy at time *t*, this is a default event. If, by contrast, Suzy does something that is incompatible with the result of our extrapolation – say, throws a rock at the window – then this is a deviant event.

To make this suggestion more precise, we first need to capture what it means to say that one event is incompatible with another. To achieve the required generality, we need to define this not just for instantaneous events, but also for temporally extended events built up from sequences of instantaneous events (as set out in Chapter 4 section 3.2). This may be defined as follows:

> *Incompatibility*: the instantaneous or temporally extended events e and e^* are *incompatible* iff $w(e) \cap w(e^*) = \varnothing$.

In other words, two events e and e^* are incompatible just in case there is no possible world where both e and e^* occur.

To begin with a simple example, let *Suzy-talks* be the event that essentially occurs within the closed interval $[t, t+dt]$, and that is essentially Suzy's standing and chatting with Billy (without throwing any rocks). And let *Suzy-throws* be the event that essentially occurs within the closed interval $[t, t+dt]$, and that is essentially Suzy's throwing a rock at the window. Then we find that *Suzy-talks* is incompatible with *Suzy-throws*.

Or, to take a more complex example, let *Suzy-keeps-talking* be the temporally extended event built up from a sequence of Suzy-talking-with-Billy events spanning the interval from t to t + 60 seconds, and let *Suzy-throws* be the event that essentially occurs within the closed interval [t, t + 60 seconds], and that is essentially Suzy's throwing a rock at the window. In this case, we now find that *Suzy-keeps-talking* is incompatible with *Suzy-throws*.

Based on this understanding of incompatibility, we may now give the following definition of what it takes for an actual instantaneous event to be a deviant event:

> *Deviant event*: an actual instantaneous event e is a *deviant* event iff there is an instantaneous or temporally extended event e^*, such that e and e^* are incompatible, and some suitable extrapolation yields the result that e^* occurs.

An actual instantaneous event e is then a *default* event just in case it is not a deviant event.

To make the definition more informative, we need to say more about what counts as a suitable extrapolation. As I said above, extrapolation consists in identifying a pattern that we then extend. To capture in more detail what it takes for an extrapolation to be suitable, it will be useful to have a name for the pattern on which a given extrapolation is based – in the following, I will call it the *extrapolation basis*. Once we are presented with a suitable extrapolation basis, we usually – as a descriptive fact – find that there is a single appropriate way (or a narrow range of appropriate ways) to extend this pattern forwards, given the practices of induction that we in fact have. The key question, therefore, is: what does it take for something to be a suitable extrapolation basis?

My suggestion is that this is determined partly by considerations concerning what it takes to be an extrapolation basis at all, and partly by the context of a particular causal inquiry:

Regarding what it takes to be an extrapolation basis at all, the following criterion needs to be satisfied: an extrapolation basis must contain a pattern that can reasonably be extended to the case at hand. In practice, this narrows down the range of possible extrapolation bases to just a few options: what is happening immediately prior to *t*, what is happening in the immediate surroundings at *t*, habitual behaviour at earlier times, typical behaviour for entities of a particular kind, and perhaps one or two more. In addition, we may in some cases take normative considerations – about what persons or entities *should* do in this situation,[19] or about what the law requires – as our extrapolation basis. But there really is not much else that can satisfy the criterion of furnishing a pattern that can reasonably be extended to the case at hand.

The context of the causal inquiry then comes into play in choosing between extrapolation bases that satisfy this basic requirement. Unfortunately, I cannot give a full answer to the question of *how* context determines which of these extrapolation bases count as 'suitable'. Instead, I will limit myself to the following observations:

---

[19] I include entities here since normative considerations sometimes apply to entities as well. For example, it is part of the proper functioning of a fire alarm that it goes off in the event of fire – and in some contexts, we may take this normative consideration as our extrapolation basis.

First, our purpose in some contexts is the attribution of praise and blame, or of legal responsibility. In such contexts, we tend to take normative considerations as our extrapolation basis – and to reject extrapolation bases that do not incorporate such normative considerations. For example, Schaffer notes the legal importance of contrasting what in fact happened with what would have happened if the defendant had acted lawfully'[20]: we here take normative considerations as our extrapolation basis and use them to generate alternatives to what in fact happened. However, when our causal inquiry is naturalistic – i.e., when we are *not* concerned with praise, blame, etc., but rather with the scientific study of what caused what – we tend to reject extrapolation bases that incorporate normative considerations, and focus instead on extrapolation bases such as what was happening just before *t*, what is happening in the immediate surroundings at *t*, etc. In the following, I will mostly focus on such naturalistic causal inquiries.

Second, considering naturalistic causal inquiries, we may in some contexts accept only the most salient extrapolation bases. Typically, this means that the *only* extrapolation basis we accept is what was happening just before *t*. In other contexts, however, we may allow less salient extrapolation bases to count as suitable, alongside the extrapolation basis consisting in what was happening just before *t*.

In the following, I will now consider examples of how my characterisation of the default-deviant distinction applies in real-life cases. We have already seen a case where an event is straightforwardly categorised as a deviant event – namely, the case of Suzy's throwing a rock at the window. Next, let us consider a case that illustrates how an event may be straightforwardly categorised as a default event:

[20] Schaffer (2010), p. 260. See also Hart and Honoré (1985); Hitchcock and Knobe (2009).

> *The prairie*: at time t, a particular little patch of land is covered in prairie grasses. All around it, the prairie stretches for miles and miles, and all this land has been covered in prairie grasses for hundreds of years.

Let *Prairie-grasses* be the event that essentially occurs within the interval $[t, t+dt]$, and that is essentially this particular little patch of land being covered in prairie grasses. By the above definition, we now find that *Prairie-grasses* is a default event: whether we extrapolate from what is happening just before t, or from the surroundings, or from any other reasonable extrapolation basis, our extrapolation yields a sequence of events from time t onwards, each of which is essentially this little patch of land being covered in prairie grasses. In this case, then, extrapolation does not yield any alternative to *Prairie-grasses* – and thus we find, as we should, that *Prairie-grasses* is a default event.

Finally, whether or not a given event is categorised as default or deviant may in some cases depend on how liberal we are regarding what counts as a suitable extrapolation basis. This is illustrated by the following case:

> *Train tracks*: at time t, there are train tracks connecting one little town with another. The tracks run through an arid landscape, and have been there for the past several years.

If we are very strict about what we accept as a suitable extrapolation basis, we may count what is happening just before t as the *only* suitable extrapolation basis. In that case, extrapolation yields no alternative to the presence of the tracks: when we extrapolate from what is happening just before t, the result of our extrapolation is that the train tracks are exactly where they in fact are. In this case, then, the event that essentially occurs within the interval $[t, t+dt]$, and that is essentially the presence of the tracks, is categorised as a default event.

If, on the other hand, we are willing to consider a wider range of extrapolation bases, we do find an alternative to the presence of the tracks: when we extrapolate from what is happening in the immediate surroundings at *t*, the result of our extrapolation is that there is nothing but air and dirt where the tracks in fact are.[21] In this case, then, the presence of the tracks is categorised as a deviant event.

Finally, it is worth noting how the default-deviant distinction applies in the case of neuron diagrams. When a neuron diagram is used to represent the causal structure of a particular real-life situation, we should of course defer to the real-life situation when determining whether a given neuron firing (or non-firing) event should be categorised as default or deviant: if it represents a default event, it should be categorised as default; if it represents a deviant event, it should be categorised as deviant.

In some cases, however, we consider neuron diagrams just as they are – as representations of a physical system of neurons that send stimulatory and inhibitory signals to each other. In such cases, we find that neuron firing events are always categorised as deviant events: neuron firings are instantaneous; when we extrapolate from what was happening earlier, we therefore always find the alternative that the neuron in question could have remained in dormant.

By contrast, we may categorise non-firing events either as default or deviant: if we only accept what was happening just before the relevant time *t* as a suitable extrapolation basis, we find no alternative to the neuron in question remaining dormant. However, if we accept other extrapolation bases – for example, the behaviour of other neurons – we may categorise a neuron's failure to fire as a deviant event.

This fits with the standard treatment of neuron firing events, where neuron firings are always categorised as deviant, while failures to fire are typically, but not always, categorised as default.[22]

---

[21] This fits nicely with the following quotation from Hall, where he asks the reader to 'subtract' the train tracks away, and adds: 'you need not go as far as to replace them by pure void; air and dirt will do' (Hall (2004*a*), p. 188).

[22] Cf. Hall (2007*b*), p. 46.

3. How context selects a possibility horizon

With the above non-causal characterisation of the default-deviant distinction, we are now ready to answer the question: how does the context of a causal inquiry select a possibility horizon?

In brief, my answer is that it does so on the basis of the default-deviant distinction. In this way, the context-sensitivity of causal claims ultimately derives from the fact that it is dependent on context which extrapolation bases count as appropriate: the selection of appropriate extrapolation bases underlies the categorisation of events as default or deviant; this categorisation in turn underlies the selection of a possibility horizon; and finally, the selected possibility horizon matters for the truth-value of our causal claims. In the remainder of this section, we shall now see in more detail how a possibility horizon is selected on the basis of the default-deviant distinction.

Let us begin by considering a simple causal claim of the form '*c* causes *e*', where *c* and *e* are instantaneous events. Let *t* be the time at which *c* occurs. My suggestion is that the default-deviant distinction applied to the instantaneous events occurring at time *t* determines the relevant possibility horizon: the context of a causal inquiry determines which of the events occurring at *t* are default events, and which of the events occurring at *t* are deviant events. This in turn places restrictions on which nomologically possible worlds may be included in the relevant possibility horizon $\mathscr{H}$:

If an actual instantaneous event *a*, which occurs at *t*, is categorised as default, this imposes the following restriction: a world *w* may be included in $\mathscr{H}$ only if *a* occurs in *w*. To illustrate, consider for example our standard case of causation by omission:

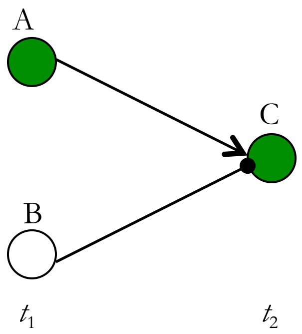


Figure 1

The class of worlds below includes all nomologically possible worlds based on this case:

	Complete state at time t_1:	*Events at t_1:*	t_2:
@:	**A** fires, **B** does not fire	A, $\neg B$	C
w_1:	**A** does not fire, **B** does not fire	A, $\neg B$	C
w_2:	**A** fires, **B** fires	A, $\neg B$	C
w_3:	**A** does not fire, **B** fires	A, $\neg B$	C

Suppose that the causal claim under consideration is 'A is a cause of C'. Thus, t_1 is selected as the relevant time. And suppose further that the context of our causal inquiry categorises $\neg B$ as a default event. In that case, the following restriction applies to our possibility horizon: a world w may be included in $\mathscr{H}$ only if $\neg B$ occurs in w. It is easy to see that there are only two worlds that satisfy this restriction, namely:

	Complete state at time t_1:	*Events at t_1:*	t_2:
@:	**A** fires, **B** does not fire	A, $\neg B$	C
w_1:	**A** does not fire, **B** does not fire	A, $\neg B$	C

If an actual instantaneous event a is categorised as a deviant event, the appropriate restriction is slightly more sophisticated. When an actual instantaneous event a is categorised as deviant, this is so because extrapolation yields at least one instantaneous or temporally extended event a^*, such that a and a^* are incompatible. Intuitively, these instantaneous or temporally extended events represent *alternatives* to a. Let $\mathbf{S} = \{a, a^*, a^{**}, \ldots\}$ be the set of events containing a together with the instantaneous or temporally extended events that extrapolation yields as alternatives to a. Then I suggest that the following restriction is appropriate: a world w may be included in $\mathscr{H}$ only if at least one event from $\mathbf{S} = \{a, a^*, a^{**}, \ldots\}$ occurs in w.

To illustrate, consider the neuron diagram below:

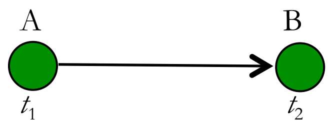


Figure 2

In addition to being in a state where it fires, neuron **A** also has the possibility of being in two states in which it does not send any outgoing signals: it may be dormant as illustrated below,

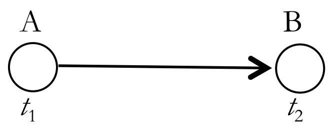


Figure 2*

or it may fire in waves:

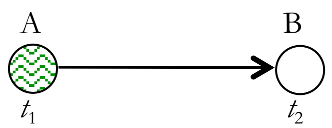


Figure 2**

Considering only maximally temporally fragile events, let A be the event that is essentially **A**'s firing, let *A-dormant* be the event that is essentially **A**'s being dormant, and let *A-waves* be the event is essentially **A**'s firing in waves.

With no restrictions, we have the following class of nomologically possible worlds characterised by their complete state at time t_1:

	Complete state at time t_1:	*Events at t_1:*			*t_2:*
@:	**A** fires	*A*,	*A-dormant*,	*A-waves*	*B*
w_1:	**A** is dormant	*A*,	*A-dormant*,	*A-waves*	*B*
w_2:	**A** fires in waves	*A*,	*A-dormant*,	*A-waves*	*B*

Suppose that the causal claim under consideration is '*A* is a cause of *B*'. Thus, the selected time is t_1. And suppose further that extrapolation yields the result that **A** is simply dormant at time t_1. In that case, *A* is categorised as a deviant event with *A-dormant* as its alternative. Thus, the appropriate restriction is that a world w may be included in $\mathscr{H}$ only if at least one of the events in **S** = {*A*, *A-dormant*} occurs in w. This winnows down the relevant class of worlds to:

	Complete state at time t_1:	*Events at t_1:*			*t_2:*
@:	**A** fires	*A*,	*A-dormant*,	*A-waves*	*B*
w_1:	**A** is dormant	*A*,	*A-dormant*,	*A-waves*	*B*

In the case of binary causal claims of the form '*c* causes *e*', my suggestion is that the possibility horizon selected by context is simply the largest class of nomologically possible worlds that satisfies all of these restrictions.

The surface form of our causal claims is usually binary. In some cases, however, we make contrastive causal claims – adding a contrast to the cause, to the effect, or both. Such contrastive causal claims take the form '*c* rather than *c** causes *e*', '*c* causes *e* rather than *e**', or '*c* rather than *c** causes *e* rather than *e**'. How does context select the relevant possibility horizon in the case of such contrastive claims?

My suggestion is that the selection in this case proceeds in two steps. First, context imposes restrictions based on *c*'s time of occurrence, precisely as set out above. Next, the specified contrasts impose *further* requirements on the relevant possibility horizon: the contrastive causal claim '*c* rather than *c** is a cause of *e*' imposes the requirements that our possibility horizon must contain at least one world in which *c* occurs, at least one world in which *c** occurs, and no world in which neither *c* nor *c** occurs. Similarly, the contrastive causal claim '*c* causes *e* rather than *e**' imposes the requirements that our possibility horizon must contain at least one world in which *e* occurs, at least one world in which *e** occurs, and no world in which neither *e* nor *e** occurs. And a quaternary contrastive claim, '*c* rather than *c** is a cause of *e* rather than *e**' obviously imposes both of these sets of requirements at once.

In some cases, there may be a conflict between the restrictions imposed by context and the requirements imposed by a particular specification of contrasts. Consider, for example, the contrastive causal claim 'the queen's reigning on her throne rather than watering Suzy's flowers caused them to die'. And suppose that context treats the queen's failure to water Suzy's flowers as a default event. This imposes the restriction that a world *w* may belong to $\mathscr{H}$ only if the queen does not water Suzy's flowers in *w*. In that case, one of the requirements imposed by the contrastive claim – namely, that our possibility horizon must contain at least one world in which the queen waters Suzy's flowers – conflicts with the contextually imposed restrictions. When this happens, the contrastive causal claim misfires: the selected contrast is irrelevant. As Schaffer writes:

> 'We resist taking such an unrealistic supposition as a contrast. The queen's watering my flowers is not easily swallowed as a *relevant alternative*. At *c** sits an irrelevance.'[23]

I return to the issue of contrastive causal claims in Chapter 10 section 3.

In the following, we shall now see how a possibility horizon selected in this way perfectly models the intuitive distinction between candidate causes and background conditions.

[23] Schaffer (2005), p. 302.

4. Candidate causes and background conditions

In our ordinary thinking about causation, we tend to distinguish between genuine causes and mere background conditions. As Hart and Honoré note:

> 'In most cases where a fire has broken out the lawyer, the historian, and the plain man would refuse to say that the cause of the fire was the presence of oxygen, though no fire would have occurred without it: they would reserve the title of cause for something of the order of a short-circuit, the dropping of a lighted cigarette, or lightning.'[24]

This distinction between causes and background conditions is clearly context-dependent, as Hart and Honoré go on to note:

> 'Yet there are contexts where it would be natural to say that the presence of the oxygen was the cause of the fire. We have only to consider a factory where delicate manufacturing processes are carried on, requiring the exclusion of oxygen, to make it perfectly sensible to identify as the cause of a fire the presence of oxygen introduced by someone's mistake.'[25]

Perhaps because of this context-dependence, the distinction between causes and background conditions has often been dismissed. For example, Mill wrote:

> 'Nothing can better show the absence of any scientific ground for the distinction between the cause of a phenomenon and its conditions, than the capricious manner in which we select from among the conditions that which we choose to denominate the cause.'[26]

And we find the same sentiment in the quotation from Lewis cited at the beginning of this chapter. As Hart and Honoré point out, however, we miss

[24] Hart and Honoré (1985), p. 11.
[25] Hart and Honoré (1985), p. 11.
[26] Mill (1973), p. 329. (Book III Chapter V section III).

something important if we dismiss the distinction between causes and background conditions. As they note:

> 'it is plain that our choice, though responsive to the varying context of the particular occasions, is not arbitrary or haphazard.'[27]

In support of this, Schaffer notes that our way of drawing the distinction between causes and background conditions is remarkably predictable. As he writes, this is 'the sort of stable intuition that philosophers normally treat as data rather than rubbish.'[28] Based on this, I believe that it would be an important advantage of the notion of a possibility horizon if it could illuminate the distinction between causes and background conditions. It can do just that:

As we have seen above, my recipe for how context selects the relevant possibility horizon ensures that whenever an event is categorised as a default event, its occurrence is *held fixed* throughout the relevant possibility horizon. Consider, for example, our standard case of causation by omission:

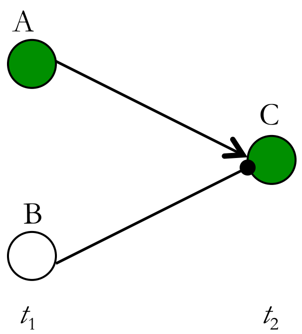


Figure 1

If $\neg B$ is categorised as a default event, we have already seen in section 3 that we get the following class of worlds characterised by their complete state at time t_1:

[27] Hart and Honoré (1985), p. 11.

[28] Schaffer (2005), p. 313. See also Menzies (2004), pp. 142-45.

	Complete state at time t_1:	*Events at t_1:*	t_2:
@:	**A** fires, **B** does not fire	A, $\neg B$	C
w_1:	**A** does not fire, **B** does not fire	A, $\neg B$	C

The occurrence of $\neg B$ is thus held fixed throughout our possibility horizon. In this way, we are essentially treating default events as background conditions.

On the other hand, deviant events – that are contrasted with their corresponding alternatives – are treated as candidate causes (of course, they then have to satisfy the conditions for causation in order to count as genuine causes). For example, A is treated as a candidate cause: it occurs in some worlds within our possibility horizon (namely in @), but fails to occur in others (namely in w_1).

The suggestion that a possibility horizon models the distinction between candidate causes and background conditions by holding the occurrence of default events fixed corresponds closely to Schaffer's characterisation of the distinction, where he suggests that:

> 'the setting of the background may be confirmed by what lacks relevant alternatives. The background is what is held fixed.'[29]

To illustrate the proposal, I will now show how this way of modelling the distinction correctly separates candidate causes from background conditions in the two kinds of cases discussed by Hart and Honoré.

As an example of a case where we would 'refuse to say that the cause of the fire was the presence of oxygen',[30] suppose that there is a lightning strike in a forest at time t_1, and a forest fire at time t_2. To simplify, let us consider just two maximally temporally fragile events occurring at time t_1 – namely, O and L, where it is essential to the occurrence of O that oxygen is present, and essential to the occurrence of L that there is a lightning strike. On any reasonable

[29] Schaffer (2005), p. 319.

[30] Hart and Honoré (1985), p. 11.

extrapolation, we do not find an alternative to *O* – whether we extrapolate from what was happening just before t_1, or from the immediate surroundings at t_1, we find that oxygen is present at t_1 just as it actually is.[31] Thus, *O* should be categorised as a default event. By contrast, we straightforwardly find an alternative to *L* – namely, the absence of abnormal electrical currents in the area in question. Thus, *L* should be categorised as a deviant event.

Based on this, we find that the relevant possibility horizon includes just the following two worlds, characterised by their complete state at time t_1:

	Complete state at time t_1:	*Events at* t_1:	t_2:
@:	presence of oxygen, lightning strike	*O*, *L*	*Forest fire*
w_1:	presence of oxygen, no lightning strike	*O*, *L*	*Forest fire*

Within this possibility horizon, the presence of oxygen is thus treated as a background condition, whereas the lightning strike is treated as a candidate cause – yielding the intuitively correct way of drawing the distinction.

Next, consider Hart and Honoré's second case: a factory where delicate manufacturing processes are carried out that require the exclusion of oxygen, and where oxygen is introduced at time t_1 by someone's mistake. To simplify, we may once again consider just two maximally temporally fragile events occurring at time t_1 – namely, *O* and *S*, where it is essential to the occurrence of *O* that oxygen is present, and where it is essential to the occurrence of *S* that a spark is present. Extrapolation straightforwardly yields an alternative to the occurrence of *O*: extrapolating from what was the case just before t_1 (or from the typical conditions of the relevant area, or from normative considerations concerning the function of the manufacturing area), we find that there is no oxygen in the relevant area at time t_1 (or, more realistically, that the

[31] In very special contexts, however, we might extrapolate from the conditions in outer space, yielding an alternative to the presence of oxygen – namely, that earth's atmosphere is absent. This explains how the claim that 'the presence of oxygen caused the forest fire' could be appropriate in a conversation among visitors from Venus (the example comes from Schaffer (2012), p. 37).

concentration of oxygen is below some very low threshold). Thus, O is categorised as a deviant event, with the alternative being that there is an appropriately low concentration of oxygen.

Regarding the presence of the spark, we may – depending on so far unspecified details of the case – either find that every reasonable extrapolation yields the result that there is a spark, or that some extrapolation yields an alternative. Assume that some reasonable extrapolation yields an alternative – for example, that all elements involved in the manufacturing process are below some threshold temperature at t_1. Thus, S is also treated as a deviant event, with all elements being below the threshold temperature as its alternative.

We thus find that our possibility horizon includes the following four worlds, characterised by their complete state at time t_1:

	Complete state at time t_1:	*Events at t_1:*	*t_2:*
@:	oxygen present, spark	O, S	*Fire*
w_1:	oxygen present, below threshold temperature	O, S	*Fire*
w_2:	no oxygen, spark	O, S	*Fire*
w_3:	no oxygen, below threshold temperature	O, S	*Fire*

In this case, then, we find – as we should – that the presence of oxygen is *not* treated as a mere background condition. Instead, both the spark and the presence of oxygen are treated as candidate causes.

In this way, the notion of a possibility horizon successfully models the intuitive distinction between causes and background conditions.

5. Conclusion

In this chapter, I have introduced the notion of a possibility horizon and shown how the context of a causal inquiry selects a possibility horizon. To capture this, I have first given a non-causal characterisation of the default-deviant distinction. I have then shown how context selects a possibility horizon based on this distinction. And finally I have shown how such a contextually selected possibility horizon captures the intuitive distinction between candidate causes and background conditions.

With this understanding of the causal relata, we are now ready to define the two necessary and jointly sufficient conditions for causation – namely, process-connection and security-dependence. It is important to note that the relation of process-connection is a binary relation: *c* is process-connected to *e*. This relation only involves the primary causal relata *c* and *e*. Thus, our choice of possibility horizon has *no* bearing on whether an event *c* is process-connected to a later event *e*. Since process-connection is a necessary condition for causation, this means that there are certain causal claims that are true independently of context: if *c* is not process-connected to *e*, then the claim that '*c* does not cause *e*' is true entirely independently of our contextually determined choice of possibility horizon.

Among the events that do satisfy the condition of process-connection, however, our choice of possibility horizon does matter: to count as causes, these events have to satisfy the further necessary condition of security-dependence. Security-dependence is a ternary relation: *e* security-depends on *c* within possibility horizon $\mathscr{H}$. Thus, an event *c* may satisfy the necessary condition of security-dependence within one possibility horizon $\mathscr{H}_1$, but fail to satisfy it within a different possibility horizon $\mathscr{H}_2$. And since security-dependence is a necessary condition for causation, this accounts for the context-dependence of our causal claims, since different contexts may select different possibility horizons.

PART III

Process-connection

6

Defining process-connection

A cause must be connected to its effect via the right kind of process. Consider, for example, a standard case of late preemption:

> *Late preemption*: Suzy and Billy both throw rocks at a window. Suzy's rock hits the window a moment before Billy's and the window shatters.

Intuitively, Suzy's throw is a cause of the shattering, while Billy's is not: Suzy's throw is a preempting cause; Billy's throw is a preempted backup. What distinguishes Suzy's throw from Billy's in relation to the window-shattering? My suggestion is that the crucial difference between them is this: Suzy's throw is – via the flight of her rock and its impact on the window-pane – connected to the window-shattering by the right kind of process. By contrast, Billy's throw is not connected to the window-shattering by the right kind of process – indeed, his rock has not even reached the window when it begins to shatter.

To develop this suggestion we need to spell out, in a suitably reductive way, what it takes for one event to be connected to another *by the right kind of process*. Doing so will be the focus of this chapter, leading up to a definition of the relation of *process-connection*, where an event *c* is process-connected to a later event *e* just in case there is a *genuine process* from *c* to *e*. My suggestion is that this relation captures a necessary condition for causation: an event *c* is a cause of a later event *e* only if *c* is process-connected to *e*.

In the following, I build up towards a reductive definition of process-connection. I begin by outlining the form that the characterisation of a process must take (section 1). Based on this, our question becomes: what does it take for a characterisation of this form to characterise a genuine process? To answer this question, I first define the relations of sufficiency and minimal sufficiency (section 2). I then build on these definitions to define the relation of time-sensitive sufficiency, which allows us to handle simple cases of late preemption (section 3). And finally, I define the notions of an apparent process and a genuine process, and set out the resulting condition of process-connection (section 4).

It is important to note that this whole discussion is conducted at the level of representations: rather than asking what conditions something must satisfy in order to be a genuine process, I ask instead what conditions a *characterisation* must satisfy in order to *characterise* a genuine process. In section 5, I discuss the advantages of ascending to the level of representations in this way.

In the following, I focus, for the sake of simplicity, on defining what it takes for something to characterise an *actual* genuine process. Since any world *w* may play the role of the actual world, this in fact provides a fully general definition of what it takes to characterise a genuine process occurring in an arbitrary world *w*.[1] However, focusing on the actual world gives a simpler statement of the definition. Note that I will in the following be talking exclusively about *actual instantaneous events*. For the sake of simplicity, I use 'event' as shorthand for 'actual instantaneous event'.

1. Characterising a process

How may we characterise a process? My suggestion is that we may do so simply by specifying a series of events. To illustrate this, consider our standard case of early preemption:

[1] Cf. Lewis (1986*d*), p. 163.

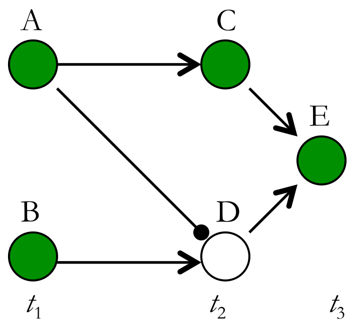


Figure 1

It is here intuitively clear that there is a process from A to E. We may characterise this process by the following series of events:

A
C
E

There is a clear sense in which such a characterisation is based on a particular time-series T given by the set of times $\{t_1, t_2, t_3\}$, since these are the times at which the events in the characterisation occur. The times t_1, t_2, and t_3 satisfy the ordering $t_1 < t_2 < t_3$, and I shall use the convention of simply writing such a time series on the form:

T: $t_1 < t_2 < t_3$

In the following, I usually make the time associated with each event explicit, as illustrated below:

t_1: A
t_2: C
t_3: E

Strictly speaking, however, this is redundant, since the relevant time can simply be read off from the quintuple characterising each event. For example, our event A is based on the quintuple $(@, t_1, s_1, I_A, C_A)$, according to our conventions for neuron events (see Chapter 4 section 2.3). From this, it is immediately clear that A occurs at time t_1. For this reason, the characterisation proper of a process should be thought of simply as a series of events. When I add a left-hand column with the relevant times, this is simply for convenience.

Obviously, not every series of events characterises a genuine process. For example, the series,

t_1: B
t_2: C
t_3: E

does not represent stages in anything we would intuitively recognise as a genuine process. Our task in the following is to answer the question: what does it take for a series of events to characterise a genuine process?

Looking at the above example, we get a hint as to how we should answer this question: the reason why we judge that there is no genuine process characterised by the series B, C, E, is that B does not stand in the right relation to C. Our first task therefore is to capture the relation that has to hold, in a genuine process, from one event to the next. To capture this relation, I will take what might look like a detour and look at the relation that has to hold between a set S_1 of contemporaneous events and a later event e_2. Once we capture this relation, I will then propose that an event e_1 stands in the right relation to a later event e_2 just in case there is a set S_1 of contemporaneous events, such that e_1 belongs to S_1 and S_1 stands in the right relation to e_2. In the following two sections, I therefore focus on the relation that has to hold between a set S_1 of contemporaneous events and a later event e_2.

2. Sufficiency and minimal sufficiency

As a first step towards capturing the relation that has to hold between a set S_1 of contemporaneous events and a later event e_2, I define what it takes for a set

S_1 of contemporaneous events to be ***sufficient*** for an event e_2. This relation makes the following two requirements:

The first requirement concerns the temporal order of occurrence of the events in S_1 and the event e_2. Since S_1 is a set of *contemporaneous* events, all the events in S_1 occur at the same time.[2] Letting t_1 be the time at which all the events in S_1 occur, we may – with a slight abuse of language – say that 'S_1 occurs at time t_1'. Based on this, the first requirement is that it must be the case that S_1 occurs strictly *earlier* than the event e_2.[3]

The second requirement is that the occurrence of all the events in S_1 must in some sense *guarantee* the occurrence of e_2. More precisely, it must be *nomologically necessary* that if all the events in S_1 occur, then e_2 also occurs.

Drawing these requirements together yields the following definition of sufficiency:

> *Sufficiency*: a set S_1 of contemporaneous events is *sufficient* for an event e_2 iff
>
> a) S_1 occurs strictly earlier than e_2, and
>
> b) for every nomologically possible world w, it is the case that if every event in S_1 occurs in w, then e_2 occurs in w.

The above definition of sufficiency is a good starting point for characterising the relation that has to hold between a set S_1 of contemporaneous events and a subsequent event e_2. However, sufficiency does not give us exactly what we need, since the above definition of sufficiency may be satisfied even when the set S_1 includes entirely irrelevant events.

To see this, suppose that the set S_1 of contemporaneous events is sufficient for a subsequent event e_2. And suppose further that the set S_1^+

---

[2] Whenever I talk about simultaneity, this is relative to a frame of reference. For the sake of simplicity, I leave relativisation to a frame of reference implicit in what follows.

[3] As mentioned in Chapter 1, my account is only intended to apply to cases where the candidate cause c occurs strictly earlier than the effect e (within an appropriately chosen frame of reference). In such cases, this requirement is unproblematic.

includes all the events in S_1, plus one or more events that occur at the same time as the events in S_1, but which are entirely irrelevant to the occurrence of e_2. In that case, it follows from the above definition that S_1^+ is also sufficient for e_2. By hypothesis, however, S_1^+ contains one or more events that are entirely irrelevant to the occurrence of e_2. We do not want to say that these entirely irrelevant events stand in the right relation to e_2. For this reason, we need to introduce the more demanding relation of *minimal sufficiency*.

Minimal sufficiency is standardly defined in terms of subset relations. For example, Paul and Hall take a set S of contemporaneous events to be minimally sufficient for a later event e just in case S is sufficient for e and no proper subset of S is sufficient for e.[4] Given my conception of the causal relata, however, merely looking at subset relations is not enough to capture what we need: suppose, for example, that c^+ is a strictly more fragile version of c. In that case, the set $\{c\}$ is obviously not a proper subset of the set $\{c^+\}$. However, we may nevertheless want to say that although both $\{c\}$ and $\{c^+\}$ are sufficient for a later event e, only $\{c\}$ is minimally sufficient, whereas $\{c^+\}$ includes irrelevant details. To capture what is needed, I therefore give a slightly more complex definition of minimal sufficiency.

The first step is to ensure that our sets of contemporaneous events contain no redundancies. So far, we have done nothing to rule out sets of contemporaneous events that repeat the same event multiple times, or that contain several more and less fragile versions of the same event. To exclude sets of this kind, we may introduce the same notation for sets that we have already introduced for events:

> For any set S of events, $w(S)$ is the set of possible worlds in which *all* the events in S occur.[5]

[4] Paul and Hall (2013), p. 16.

[5] Note that this definition requires us to quantify over absolutely all metaphysically possible worlds, with no restrictions on our domain of quantification. It does not merely require us to

Based on this, we may now capture what it takes for a set of contemporaneous events to be *clean* in the required sense:

> *Clean set*: a set S of events is *clean* iff
> there is no set S^* such that
> a) S^* is a proper subset of S, and
> b) $w(S^*) = w(S)$.

This excludes redundancies of the kind mentioned above: a set of contemporaneous events that repeats the same event multiple times or contains more and less fragile versions of the same event cannot satisfy the above definition. In the following, I will be concerned exclusively with clean sets. Whenever I talk of a 'set', this should therefore be understood to mean a 'clean set' according to the above definition.

The second step in capturing the notion of minimal sufficiency is to ensure that every event in our set is indeed relevant to the effect we are interested in. To do this, we need to capture the idea that one set of events can be a more or less fragile version of another. In Chapter 4 section 2.2, we have already seen how one *event* can be a more fragile version of another. For ease of reference, I repeat the definition here:

> *More fragile version – definition 1*: an event $e^+ = (@, t^+, s^+, I^+, C^+)$
> is a *more fragile version* of an event $e = (@, t, s, I, C)$ iff
> a) e^+ and e have the same realization in @, and
> b) $w(e^+) \subseteq w(e)$.

We may now extend this to also cover *sets* of contemporaneous events: it follows from my account of the causal relata that any two events that occur at the same time have the same realization. This means that *all* the events in a set

quantify over nomologically possible worlds, or worlds that are included in the contextually relevant possibility horizon.

of contemporaneous events have the same realization. Because of this, we may simply take the realization of a set of contemporaneous events to be the realization that all its members share. Building on this, we may now define what it takes for one set of contemporaneous events to be a more fragile version of another as follows:

> *More fragile version of a set*: a set S^+ of contemporaneous events is a *more fragile version* of a set S of contemporaneous events iff
> a) S^+ and S have the same realization in @, and
> b) $w(S^+) \subseteq w(S)$.

This definition includes the limit case where S^+ is identical to S as a case where S^+ is a more fragile version of S. When a set S^+ is a more fragile version of a set S, and $w(S^+) \subset w(S)$, we may say that S^+ is a *strictly* more fragile version of S.[6]

The above definition gives us exactly what we need in order to capture the relation of minimal sufficiency:

> *Minimal sufficiency*: a set S_1 of contemporaneous events is *minimally sufficient* for an event e_2 iff
> a) S_1 is *sufficient* for e_2, and
> b) there is no set S_1' of contemporaneous events, such that
>   i) S_1 is a strictly more fragile version of S_1', and
>   ii) S_1' is *sufficient* for e_2.

This definition of minimal sufficiency brings us a step closer to capturing the relation we are after. In the following section, I will now build on this to capture a relation that can differentiate between genuine causes and preempted backups in simple cases of late preemption.

---

[6] I sometimes refer to a *less fragile* set as being *more robust*.

3. Time-sensitive sufficiency

The above definition of minimal sufficiency cannot differentiate between genuine causes and preempted backups in cases of late preemption. To see this, consider our standard case of late preemption:

> *Late preemption*: Suzy and Billy both throw rocks at a window. Suzy's rock hits the window a moment before Billy's and the window shatters.

Intuitively, Suzy's throw is a cause of the window-shattering, while Billy's is not. Correspondingly, we want to say that there is a genuine process from Suzy's throw to the window-shattering, while there is no genuine process from Billy's throw to the window-shattering.

The above notion of minimal sufficiency, however, does not allow us to draw this distinction. To see this, let us consider a more detailed specification of the case: suppose that Suzy and Billy are both standing 7 meters away from the window, and both throw their rocks 0.5 second before 12 noon. Suzy throws her rock at a speed of 14 m/s; Billy throws his rock at a speed of 10 m/s; and the window begins to shatter at precisely 12 noon. Finally, suppose that we are interested in the following window-shattering event:[7]

> *Window-shattering*: the event characterised by (@, t_2, s_2, I_w, C_w), where
> t_2 = 12 noon,
> s_2 = the complete state of @ at time t_2,
> I_w = $]-\infty$, 12.01 p.m.],
> C_w = the class of complete states such that the window begins to shatter.

---

[7] Note that this event is based on the interval $]-\infty$, 12.01 p.m.], i.e. it is essential to it that it occurs at *some* time before 12.01 p.m. (on a particular day, which I have left unspecified for the sake of simplicity). Also note that the interval $]-\infty$, 12.01 p.m.] is indeed a closed interval; the reason that it does not include $-\infty$ is that, obviously, there is no such time as $-\infty$.

Half a second before 12 noon, the following two events both occur:[8]

Suzy's throw: the event characterised by ($@$, t_1, s_1, I_S, C_S), where
t_1 = 0.5 second before 12 noon,
s_1 = the complete state of @ at time t_1,
I_S = $]-\infty$, 0.5 second before 12.01 p.m.],
C_S = the class of complete states such that the window is intact, and Suzy, standing 7 meters from the window, throws her rock towards it with a speed of 14 m/s.

Billy's throw: the event characterised by ($@$, t_1, s_1, I_B, C_B), where
t_1 = 0.5 second before 12 noon,
s_1 = the complete state of @ at time t_1,
I_B = $]-\infty$, 0.7 second before 12.01 p.m.],
C_B = the class of complete states such that the window is intact, and Billy, standing 7 meters from the window, throws his rock towards it with a speed of 10 m/s.

By our definition of minimal sufficiency, we now find that both of the following claims are true:

{*Suzy's throw*} is minimally sufficient for *Window-shattering*.
{*Billy's throw*} is minimally sufficient for *Window-shattering*.

Thus, our definition of minimal sufficiency does not capture the important difference between the way in which *Suzy's throw* is related to *Window-shattering*, and the way in which *Billy's throw* is related to *Window-shattering*. How can we strengthen the definition of minimal sufficiency in a way that allows us to capture this difference?

8 For the sake of simplicity, I give a somewhat incomplete characterisation of these two events – for example, I leave out that there are no objects blocking the path towards the window, etc.

Intuitively, the crucial difference between Suzy's throw and Billy's throw has to do with the *timing* of the window-shattering: Billy's rock arrives *too late* to be a cause of the shattering. Within the context of counterfactual theories of causation, there have been several proposals attempting to leverage this intuition into an account that can handle late preemption.[9] However, it is now recognised that these proposals all face serious problems.[10]

In the following, I propose a new way to leverage this intuition into a strengthened definition of minimal sufficiency. To do this, I will rely on my definition of what it takes for one event to be a more *temporally* fragile version of another (see Chapter 4 section 2.2). For ease of reference, I repeat the definition here:

> *More temporally fragile version*: an event $e^+ = (@, t^+, s^+, I^+, C^+)$ is a *more temporally fragile version* of an event $e = (@, t, s, I, C)$ iff
>
> a) e^+ and e have the same realization in @,
>
> b) $I^+ \subseteq I$, and
>
> c) $C^+ = C$.

Based on this definition, it is clear that there is a maximally temporally fragile version of the window-shattering, namely:

> *Precise window-shattering*: the event characterised by $(@, t_2, s_2, I_p, C_w)$, where
>
> t_2 = 12 noon,
>
> s_2 = the complete state of @ at time t_2,
>
> I_p = [12 noon, 12 noon + dt].
>
> C_w = the class of complete states such that the window begins to shatter.

A natural first suggestion is that, in order to distinguish the relation between Suzy's throw and the window-shattering from the relation between Billy's

[9] See, in particular, Paul (1998*a*) and Lewis (2004*a*).

[10] See e.g. Hall and Paul (2003), pp. 112-14 and Paul and Hall (2013), pp. 102-110.

throw and the window-shattering, we need to pay attention to this maximally temporally fragile version of the window-shattering.

However, this first suggestion does not in fact succeed in capturing the difference between the two relations: it may well be the case – indeed, it almost certainly is the case – that Billy's throw has some tiny effect on the time at which the window begins to shatter.[11] Thus, we may well find that a minimally sufficient set for *Precise window-shattering* needs to include both Suzy's throw and Billy's throw. Looking at *Precise window-shattering* therefore does not help us to capture how the relation between Suzy's throw and the shattering *differs* from the relation between Billy's throw and the shattering.

Rather than merely paying attention to the maximally temporally fragile version of the window-shattering, my suggestion therefore is that we need to pay attention to *all* the intermediate more temporally fragile versions of *Window-shattering*: it is exactly by paying attention to *intermediate* levels of detail that we may capture what distinguishes the relation between Suzy's throw and the window-shattering from the relation between Billy's throw and the window-shattering.

More precisely, I suggest that we need the following strengthened version of minimal sufficiency, which I call *time-sensitive sufficiency*:

> *Time-sensitive sufficiency*: a set S_1 of contemporaneous events is *time-sensitively sufficient* for an event e_2 iff
>
> a) S_1 is minimally sufficient for e_2, and
>
> b) for every more temporally fragile version e_2^+ of e_2, there is a set S_1^+ of contemporaneous events, such that
>
> i) S_1^+ is a more fragile version of S_1, and
>
> ii) S_1^+ is minimally sufficient for e_2^+.

[11] See e.g. Paul and Hall (2013), pp. 105-6.

This definition captures the intuition that when there really is a process out in the world, we can take a closer look, and there will still be a process: when we consider a more temporally precise specification e_2^+ of e_2, we can find a correspondingly more precise specification S_1^+ of S_1, such that the right relation still holds between S_1^+ and e_2^+.

This strengthened definition now allows us to distinguish the relation between Suzy's throw and the window-shattering from the relation between Billy's throw and the window-shattering: {*Suzy's throw*} is time-sensitively sufficient for *Window-shattering*, whereas {*Billy's throw*} is not.

To see this, consider for example the following event, which is a more temporally fragile – but not maximally temporally fragile – version of *Window-shattering*:

Window-shattering$^+$: the event characterised by $(@, t_2, s_2, I_w^+, C_w)$, where
t_2 = 12 noon,
s_2 = the complete state of @ at time t_2,
$I_w^+ =]-\infty$, 0.1 second after 12 noon],
C_w = the class of complete states such that the window begins to shatter.

We can here easily find a corresponding more fragile version, $\{$*Suzy's throw*$\}^+$, of {*Suzy's throw*}, such that $\{$*Suzy's throw*$\}^+$ is minimally sufficient for *Window-shattering*$^+$, namely the singleton set containing:

Suzy's throw$^+$: the event characterised by $(@, t_1, s_1, I_S^+, C_S)$, where
t_1 = 0.5 second before 12 noon,
s_1 = the complete state of @ at time t_1,
$I_S^+ =]-\infty$, 0.4 second before 12 noon], and
C_S = the class of complete states such that the window is intact, and Suzy, standing 7 meters from the window, throws her rock towards it with a speed of 14 m/s.

And although I cannot explicitly test every case, it is clear that we can go on to do the same for any more temporally fragile version of *Window-shattering*.

In the case of Billy's throw, on the other hand, consideration of *Window-shattering*$^{+}$ shows that {*Billy's throw*} is *not* time-sensitively sufficient for *Window-shattering*: the *only* set of contemporaneous events occurring at time t_1 that is minimally sufficient for *Window-shattering*$^{+}$ is {*Suzy's throw*$^{+}$}. But clearly, {*Suzy's throw*$^{+}$} is not a more fragile version of {*Billy's throw*}. Thus we find, as we should, that {*Billy's throw*} is *not* time-sensitively sufficient for *Window-shattering*.

I suggest that *Time-sensitive sufficiency* characterises the relation that has to hold from a set of contemporaneous events to a subsequent event.

It is worth noting that when we are dealing with neuron events (including non-firing events) defined in accordance with the conventions set out in Chapter 4 section 2.3, the relations of minimal sufficiency and time-sensitive sufficiency are co-extensive. The reason is that, according to our conventions, the neuron events we consider are already maximally temporally fragile. Thus, it follows that for any neuron event E it is the case that if a set S_1 of contemporaneous events is minimally sufficient for E, then S_1 is also time-sensitively sufficient for E.

4. Apparent processes, genuine processes, and process-connection

In sections 2 and 3 we have focused on the relation that has to hold between a set S_1 of contemporaneous events and a later event e_2, and this has led us to the relation of *Time-sensitive sufficiency*. Building on this, we may now define what it takes for a series of events to characterise a genuine process.

It will be useful to have a name for a series of events where each event belongs to a set of contemporaneous events that is time-sensitively sufficient for the next. Let us say that such a series characterises an *apparent process*, defined as follows:

> *Apparent process*: a series of events $e_1, e_2, \ldots, e_n$, based on the time-series T: $t_1 < t_2 < \ldots < t_n$, characterises an *apparent process* iff for all $i < n$, there is a set S_i of contemporaneous events, such that
>
> a) e_i belongs to S_i, and
>
> b) S_i is time-sensitively sufficient for e_{i+1}.

This notion of an apparent process comes very close to capturing what we intuitively understand to be a genuine process.[12] However, there is still something left that we need to address: so far, I have not imposed any conditions on the time-series on which our characterisation of a process is based. As I will now show, however, our choice of time-series sometimes makes a crucial difference. To see this, let us return to our standard case of early preemption:

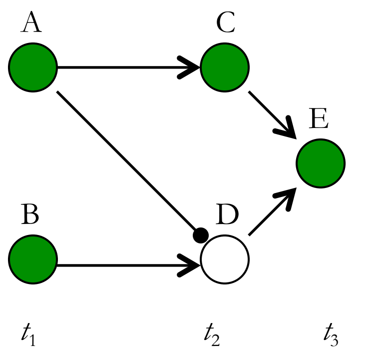


Figure 1

As we have seen, it here seems intuitively correct to say that there is a genuine process from A to E, but not from B to E. However, not all time-series allow us to draw this distinction. To see this, consider the following time-series:

> T: $t_1 < t_3$

Based on T, we find that there is a series of events that characterises an apparent process from A to E, namely:

---

[12] Note that my definition of an apparent process is structurally similar to the much simpler definition found in McDermott (1995), pp. 535-36.

P_1:

t_1: A

t_3: E

And we similarly find that there is a series of events that characterises an apparent process from B to E, namely:

P_2:

t_1: B

t_3: E

Thus, our definition of an apparent process does not differentiate between the genuine cause and the preempted backup. How can we strengthen our definition in a way that allows us to draw this distinction?

My suggestion is that we need to pay attention to what happens when we add more times to the time-series under consideration. To motivate this suggestion, consider what happens when we add one more time, namely time t_2, to the time-series under consideration, yielding the time-series:

T^+: $t_1 < t_2 < t_3$

Based on this new time-series, we still find that there is a series of events P_1^+, characterising an apparent process from A to E, namely the following:

P_1^+:

t_1: A

t_2: C

t_3: E

By contrast, there is *no* series of events based on T^+ that characterises an apparent process from B to E, since it is impossible to find an event occurring at time t_2 that allows us to satisfy the definition of an apparent process:

P_2^+:

t_1: *B*

t_2: ?

t_3: *E*

This hints at the crucial difference between P_1 and P_2. In the case of P_1, we find that our original characterisation of the process from *A* to *E* remains *stable* when we add more times to the time-series under consideration: when we add more times – in this case, when we add t_2 – yielding an augmented time-series T^+, there is a corresponding new series of events, namely P_1^+, based on T^+, such that i) P_1^+ characterises an apparent process, and ii) for every time *t* in our original time-series *T*, P_1 and P_1^+ associate *exactly the same event* with *t*. In the case of P_2, on the other hand, we find that when we consider the augmented time-series T^+, there is *no* series of events P_2^+ based on T^+, such that i) P_2^+ characterises an apparent process, and ii) for every time *t* in our original time-series *T*, P_2 and P_2^+ associate *exactly the same event* with *t*.

To leverage this observation into a definition of what it takes for a series of events to characterise a genuine process, we first need a definition of when one time-series is an augmented version of another:

> *Augmented version*: a time-series T^+ is an *augmented version* of a time-series *T*, where *T* starts from time t_1 and ends with time t_n, iff
>
> a) T^+ starts from time t_1 and ends with time t_n, and
>
> b) the times in *T* are a (proper or improper) subset of the times in T^+.

Note that it follows from this definition that in the limit case where T^+ is identical with *T*, T^+ is an augmented version of *T*.

Based on this, we may now introduce the notion of a master-set for a particular time-series:

> *Master-set*: a set **T** of time-series is the *master-set* for a time-series T iff:
> a time-series T^+ belongs to **T** iff T^+ is an augmented version of T.

More perspicuously, the master-set **T** for a given time-series T is the set containing *every* augmented version of T.

Based on this, we may now finally define what it takes for a series of events to characterise a genuine process:

> *Genuine process*: a series of events P based on a time-series T characterises a *genuine process* iff P belongs to a set **P** of series of events, such that
>
> a) there is a one-one mapping between **P** and the master-set **T** for T that maps a series of events P_i from **P** to a time-series T_i in **T** iff P_i is based on T_i,
> b) each series of events in **P** characterises an apparent process, and
> c) for any time t, all series of events in **P** that associate an event with t associate the same event with t.

This definition of a genuine process is quite demanding: it requires us to consider *every* augmented version of our time-series T.[13]

It is worth noting, however, that in the case of neuron diagrams this turns out to be a very simple task. To see this, recall our convention of treating the information contained in a neuron diagram as *complete* (see Chapter 2 section 1.1). This convention makes it very easy to reach conclusions about when an apparent process is also a genuine process: neuron diagrams only contain explicit information about what happens at a few specific times. For example, the neuron diagram illustrating our early preemption case in Figure 1

[13] However, it should be noted that this very demanding condition seems to accurately reflect the strategy we use when we are trying to verify that a series of events really represents a genuine process out in the world. This strategy is invariably: take a closer look. If by taking a closer look we find that what happens at some intermediate time does not fit with our hypothesis that there is a genuine process from one event to another, then we have to reject the hypothesis.

above only contains explicit information about what happens at time t_1, t_2, and t_3. This means that when we have a series of events characterising an apparent process based on the time-series T: $t_1 < t_2 < t_3$, this characterisation will remain stable as we add more times to the time-series under consideration – for the simple reason that the neuron diagram contains no additional information about what is happening at these intermediate times. We therefore have the following general result:

For any series of actual events P based on a time-series T, it is the case that if

a) P characterises an apparent process, and
b) T includes exactly those times at which the neuron diagram shows explicitly what the neurons do,

then P characterises a *genuine* process.

Based on this, it is easy to see that our definition of genuine processes yields the correct results when applied to our case of early preemption in Figure 1: considering the time-series,

T: $t_1 < t_2 < t_3$

we find that there is a genuine process from A to E, whereas there is no genuine process from B to E.

With this understanding of genuine processes, I am now ready to define the relation of process-connection:

Process-connection: an event *c* is *process-connected* to an event *e* iff there is a genuine process from *c* to *e*.

I propose that this relation captures a necessary condition for causation:

The first condition – process-connection:
c is a cause of *e only if c* is process-connected to *e*.

In Appendix A, I show that this relation has three important properties. First, counterfactual dependence is not sufficient for process-connection. However, given a very plausible assumption, we can prove a restricted principle of sufficiency of counterfactual dependence for process-connection (Appendix A section 1). Second, process-connection is a transitive relation (Appendix A section 2). And third, the relation of process-connection is intrinsic to a process (Appendix A section 3).

I believe that the necessary condition of process-connection correctly captures what distinguishes genuine causes from preempted backups. In our late preemption case above, for example, we judge that Billy's throw is not a cause of the window-shattering – and I suggest that we make this judgement precisely because there is no genuine process from Billy's throw to the shattering. Similarly, in our early preemption case above, we judge that B is not a cause of E – and again, I suggest that we make this judgement precisely because there is no genuine process from B to E.

5. Ascending to the level of representations

The definition of a genuine process that I have given above operates at the level of representations: rather than saying directly what it takes for something to be a genuine process, it tells us what it takes for a series of events – i.e., a *characterisation* of a process – to characterise a genuine process. This still tells us about what is out there in the world: if we find that a series P of events characterises a genuine process, then this tells us that there *is* a genuine process out in the world, which is truly (albeit incompletely) characterised by P. Even so, one might well ask: why ascend to the level of representations in this way? Why not define directly what it takes for something to be a genuine process, rather than what it takes for a series of events to characterise a genuine process?

In this section I set out my motivation for ascending to the level of representations. Surprisingly, perhaps, my motivation derives from questions about the structure of time: at present, we do not know whether time is discrete (with a structure corresponding to the natural numbers), dense (with a structure corresponding to the rational numbers), or continuous (with a structure corresponding to the real numbers) – though there are important arguments for holding that time is either dense or continuous.[14] Furthermore, it is an open question whether the structure of time is metaphysically necessary or contingent[15] – and thus, even discovering that time in the actual world is discrete would still leave it open whether there are other metaphysically possible worlds in which time is dense or continuous. In order to ensure that my proposed account of causation is applicable in any metaphysically possible world governed by deterministic laws (that satisfy the further requirements set out in Chapter 3), I therefore need to ensure that the application of my account is independent of whether time is discrete, dense, or continuous.

This, however, creates a challenge: a central feature of our intuitive understanding of what it takes for something to be a genuine process concerns the relation that has to hold between one event in a process and the next. Making sense of this notion of 'the next event' is simple if the structure of time is discrete: in this case, time itself consists in a series of discrete times (moments or instants) – t_1, t_2, t_3, . . . – and the next event in a process is simply the event associated with the next time (i.e. the immediate successor) in the series. If time is dense or continuous, however, the notion of 'the next event' runs into trouble: the only principled way to understand the notion of 'the next event' is as 'the immediately following event'. If time is dense or continuous, however, there simply *is* no immediately following event: for any two times t and t', there is always a further time t^*, such that t^* lies *between* t and t' – thereby defeating any claim that the event associated with time t' follows immediately after the event associated with t.

[14] See e.g. Hawley (2004), p. 51.

[15] See e.g. Forrest (1995); Le Poidevin (1990), p. 420; Newton-Smith (1980), pp. 112-126; Traynor (2014), pp. 183-84.

It is precisely to solve this problem that I have chosen to ascend to the level of representations: the great weakness of representations – namely the fact that, even when true, they may be *incomplete* – is in this case also their great strength. For, by ascending to the level of representations, we may characterise any process – whether time is in fact discrete, dense or continuous – based on a series of discrete times. This is exactly what I have done here, thereby giving a straightforward interpretation to the intuitive notion of 'the next event' – namely, as the event associated with the next time *within the time-series chosen in the representation.* In this way, the definition I have given works exactly by taking advantage of the fact that ascending to the level of representations allows for incompleteness.

At the same time, however, it is of course crucial to guard against the weakness of representations: the fact that even a true representation, precisely because of its incompleteness, may create the appearance that something is the case, when in fact it is not. This is what motivates my final step of moving from the definition of an apparent process to the definition of a genuine process: this final step is precisely designed to ensure that the conclusions we draw on the basis of an incomplete representation – incomplete, because based on an incomplete time-series – remain stable as we gradually move closer to a complete representation by adding more and more times to the time-series under consideration.

6. Conclusion

Intuitively, the difference between a genuine cause and a preempted backup is that a genuine cause is connected to its effect via the right kind of process, whereas a preempted backup is not. My aim in this chapter has been to leverage this intuition into a necessary condition for causation, namely the condition of *process-connection*. The challenge in developing this condition has been to give a suitably reductive characterisation of the notion of a genuine process. This then immediately leads to the condition of process-connection.

This condition, I believe, captures the intuitive thought that a cause must be *connected to its effect via a genuine process*. In the following chapter, we shall see how this condition allows us to arrive at intuitively correct verdicts in a wide range of cases.

7

Applications of process-connection

My aim in this chapter is to give an overview over the main applications of the condition of process-connection. I begin with a short section showing how the condition of process-connection allows us to distinguish causation from mere correlation (section 1). I next show how the condition of process-connection allows us to deal with three important groups of cases: cases of redundant causation (section 2), cases of omission-involving causation, including cases of omission-involving redundant causation (section 3), and cases that require commensuration between cause and effect (section 4). In all these cases, we find that an event *c* is a cause of a later event *e* just in case *c* is process-connected to *e*.

1. Correlation without causation

Let us begin by considering a case of correlation without causation. It is a common misconception that we need a counterfactual account of causation in order to handle cases of this kind. As Paul and Hall have pointed out, this misconception is due to a conflation between two different rivals to counterfactual accounts of causation: regularity accounts, which do have trouble with cases of correlation without causation, and minimal sufficiency accounts, which do not.[1]

In the following, I show how the condition of process-connection, which at its core is a minimal sufficiency condition, can easily distinguish between genuine causation and mere correlation. To see this, consider for example the case illustrated below:

[1] Paul and Hall (2013), pp. 72-73.

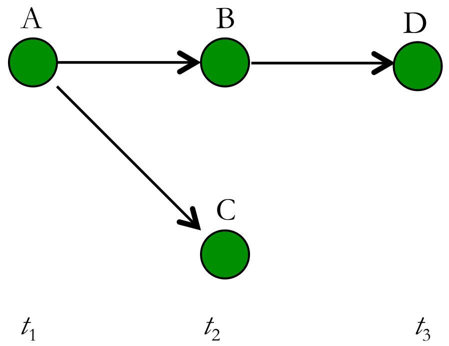


Figure 1

Within this system of neurons, D is correlated with C: whenever C occurs, so does D, since C and D have a common cause, namely A. However, C is not a cause of D: among the events occurring at time t_2, B is the only cause of D. Correspondingly, we find that B is process-connected to D, while C is not.

To see this, consider the time-series T: $t_2 < t_3$. Based on this time-series, we find that there is an apparent process from B to D, namely:

t_2: B
t_3: D

Since the time-series T: $t_2 < t_3$ includes all the times between t_2 and t_3 that are explicitly represented in the neuron diagram, this series of events characterises a *genuine* process (cf. the result on neuron diagrams presented at the end of Chapter 6 section 4). Thus, B is process-connected to D.

By contrast, there is no genuine process from C to D, since C does not belong to a minimally sufficient set for D: the set $\{C\}$ is not sufficient for D, since there is a nomologically possible world where C occurs and D does not – for example, a world that starts out at time t_2 in a complete state where **C** fires and **B** does not.[2] And the set $\{B, C\}$ is sufficient, but not minimally sufficient, since $\{B, C\}$ is a strictly more fragile version of $\{B\}$, which is already sufficient for D. Thus, we find – as we should – that C is *not* process-connected to D.

[2] Cf. Paul and Hall (2013), pp. 72-73.

2. Redundant causation

In this section, I show how the condition of process-connection enables us to successfully distinguish between genuine causes and preempted backups in standard cases of early preemption (section 2.1), late preemption (section 2.2), trumping preemption (section 2.3), and symmetric overdetermination (section 2.4).

2.1 Early preemption

Let us first consider our standard case of early preemption, illustrated below:

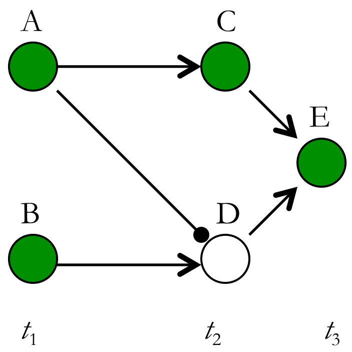


Figure 2

Intuitively, *A* is a cause of *E*, while *B* is not. The results of applying the condition of process-connection fit perfectly with these intuitive data:

The only time-series we need to consider is *T*: $t_1 < t_2 < t_3$. And the only series of events based on *T* that characterises an apparent process is the following:

t_1: *A*
t_2: *C*
t_3: *E*

Since *T* includes all the times that are explicitly represented in the neuron diagram, this is a genuine process. I have illustrated this process here, by putting a halo around the neurons that are part of the process:

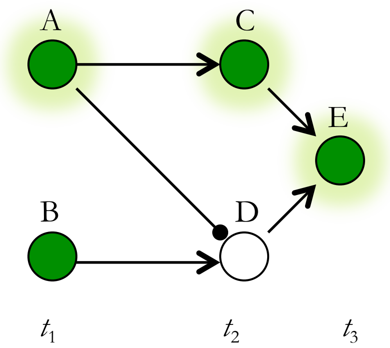


Figure 2*

From this, it is clear that A is process-connected to E, while B is not. Thus, the condition of process-connection successfully distinguishes between the genuine cause and the preempted backup in the present case.[3]

2.2 Late preemption

Let us next turn to our standard case of late preemption:

> *Late preemption*: Suzy and Billy both throw rocks at a window. Suzy's rock hits the window a moment before Billy's and the window shatters.

Intuitively, Suzy's throw is a cause of the window-shattering, while Billy's is not: Suzy's throw is a preempting cause; Billy's is a preempted backup.

This intuitive judgement fits perfectly with what we find when we apply the condition of process-connection: suppose that the details of the case are as specified in Chapter 6 section 3 and that the effect we are interested in is:

> *Window-shattering*: the event characterised by $(@, t_2, s_2, I_w, C_w)$, where
> t_2 = 12 noon,
> s_2 = the complete state of @ at time t_2,
> I_w = $]-\infty$, 12.01 p.m.],
> C_w = the class of complete states such that the window begins to shatter.

---

[3] For my treatment of a variant of early preemption, see Appendix B case 3.

As we have seen in Chapter 6 section 3, there is at the earlier time $t_1 = 0.5$ second before 12 noon, only one set of contemporaneous events that is *time-sensitively sufficient* for *Window-shattering*, namely the set {*Suzy's throw*}, where:

> *Suzy's throw*: the event characterised by (@, t_1, s_1, I_S, C_S), where
> t_1 = 0.5 second before 12 noon,
> s_1 = the complete state of @ at time t_1,
> I_S =]−∞, 0.5 second before 12.01 p.m.],
> C_S = the class of complete states such that the window is intact, and Suzy, standing 7 meters from the window, throws her rock towards it with a speed of 14 m/s.

This means that there is only one apparent process leading to *Window-shattering*, based on the time-series T: $t_1 < t_2$, namely:

> t_1: *Suzy's throw*
> t_2: *Window-shattering*

It should be obvious that this apparent process is indeed a genuine process: to any intermediate time, we can easily associate an appropriate event, specifying the relevant position and speed of Suzy's rock as it flies towards the still intact window. This yields the intuitively correct result: *Suzy's throw* is process-connected to *Window-shattering*. By contrast, let us now consider *Billy's throw*:

> *Billy's throw*: the event characterised by (@, t_1, s_1, I_B, C_B), where
> t_1 = 0.5 second before 12 noon,
> s_1 = the complete state of @ at time t_1,
> I_B =]−∞, 0.7 second before 12.01 p.m.],
> C_B = the class of complete states such that the window is intact, and Billy, standing 7 meters from the window, throws his rock towards it with a speed of 10 m/s.

Clearly, {*Billy's throw*} is minimally sufficient for *Window-shattering*. However, as we have already seen in Chapter 6 section 3, {*Billy's throw*} is not time-sensitively sufficient for *Window-shattering*. Thus, the series of events,

t_1: *Billy's throw*
t_2: *Window-shattering*

does not characterise an apparent process. Nor can it be brought to characterise an apparent process by adding intermediate times between t_1 and t_2. And so, we find – as we should – that *Billy's throw* is not process-connected to *Window-shattering*.

This shows how the condition of process-connection distinguishes between genuine causes and preempted backups in cases of late preemption.[4]

2.3 A variant of trumping preemption

As a final example of how the condition of process-connection deals with preemption, let us consider a variant of trumping preemption. Schaffer's original case of trumping preemption[5] involves action at a temporal distance and thus lies outside the scope of this dissertation. However, the following case developed by Paul and Hall presents essentially the same challenge:[6]

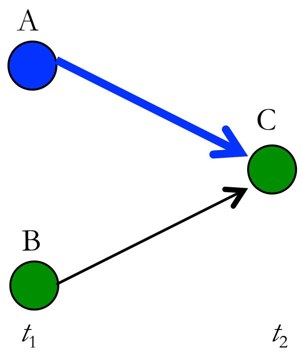


Figure 3

[4] For my treatment of a different variant of late preemption, see Appendix B case 5.
[5] See Schaffer (2004). For discussion, see also Lewis (2004*a*) and Bernstein (2015).
[6] Cf. Paul and Hall (2013), Figure 26.

Neuron **A** can here fire in different shades and in different intensities: it can fire in blue or green, and it can fire with intensity *I* or with some other intensity. Neuron **B**, on the other hand, either fires or does not fire – with no distinctions between different shades or intensities. The nomological relationships are summarised in the table below:[7]

	B fires	**B** doesn't fire
A doesn't fire	**C** fires	**C** doesn't fire
A fires in blue	**C** fires	**C** fires
A fires with intensity *I*, and not in blue	**C** doesn't fire	**C** doesn't fire
A fires without intensity *I*, and not in blue	**C** fires	**C** doesn't fire

The case we are interested in – and which is illustrated in Figure 3 above – is the case where **A** and **B** both fire, and where **A** fires in blue and with intensity *I*. On a reasonable interpretation of the case, which is also the interpretation that Paul and Hall favour, the intuitively correct verdict is that **A**'s firing in blue is a cause of **C**'s firing, and that **B**'s firing is *not* a cause of **C**'s firing.[8]

In support of this, note that this verdict is simply a generalisation of the patterns we can observe in other scenarios: as can be seen from the table above, **A**'s firing in blue is followed by **C**'s firing, even when **B** does not fire – making the name 'triggering' blue appropriate. And furthermore, when **A** fires with intensity *I*, **C** does not fire – even when **B** fires – making the name 'inhibiting' intensity appropriate. This lends support to the suggested interpretation of the case. As Paul and Hall note:

> 'What would seem most natural […] is to project the causal structures appropriate for describing the cases in general onto the test case. If so, we

[7] Cf. Paul and Hall (2013), p. 136.

[8] Paul and Hall (2013), p. 137.

> should insist that the right interpretation of the test case tells us (i) that the *A*-signal prevents the *B*-signal from causing *C* (because the *A*-signal has the inhibiting intensity *I*); and (ii) that the *A*-signal itself causes *C* (because it is in the triggering shade […]). Hence, in this neuron world, it is *A* and not *B* that causes *E*.'[9]

As Paul and Hall note, the structure of this case comes very close to the structure of Schaffer's original case of trumping preemption. Can my condition of process-connection deliver the intuitively correct result in this case?

I believe it can. Let the effect we are interested in be **C**'s firing – i.e., the event *C*. We now find that there is exactly one set of events occurring at time t_1 that is minimally sufficient for *C*, namely the set {*A-blue*}, where the instantaneous event *A-blue* is defined as follows:

> *A-blue*: the instantaneous event based on $(@, t_1, s_1, I_A, C_A)$, where
> $I_A = [t_1, t_1 + dt]$, and
> C_A = the class of states such that neuron **A** fires in blue.

From the fact that {*A-blue*} is minimally sufficient for *C*, it immediately follows that there is a genuine process from *A-blue* to *C*, namely:

> t_1: *A-blue*
> t_2: *C*

Thus, we find – exactly as we should – that *A-blue* is process-connected to *C*.

We similarly find – again, exactly as we should – that *B* is *not* process-connected to *C*, since *B* does not belong to a set of contemporaneous events that is minimally sufficient for *C*. To see this, note that {*B*} is not sufficient for *C*, since there is a nomologically possible world in which *B* occurs and *C* does not occur – namely, a world in which **A** fires with intensity *I*, without firing in

[9] Paul and Hall (2013), p. 137.

blue. And the set $\{B, A\text{-}blue\}$, though sufficient, is not minimally sufficient, since $\{B, A\text{-}blue\}$ is a strictly more fragile version of $\{A\text{-}blue\}$, which is itself sufficient for C. Thus, we find that B is not process-connected to C.

2.4 Symmetric overdetermination

I turn next to cases of symmetric overdetermination. Like preemption cases, such cases have created a good deal of trouble for counterfactual accounts of causation, since the effect in a case of symmetric overdetermination does not depend counterfactually on either of its causes. However, the causes in a case of symmetric overdetermination easily satisfy the condition of process-connection. To see this, consider the simple case illustrated below:[10]

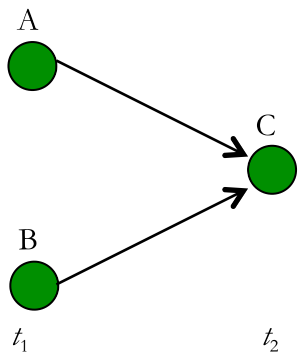


Figure 4

Here, neuron **C** fires if and only if it receives at least one stimulatory signal. Furthermore, **C** fires at exactly the same time and in exactly the same manner no matter whether it receives one or two stimulatory signals.

Intuitively, A is a cause of C, and by symmetry, so is B. Correspondingly, we find that both satisfy the necessary condition of process-connection. To see this, we simply need to consider the time-series T: $t_1 < t_2$. We now find that there is a genuine process from A to C, namely:

t_1: A
t_2: C

[10] Cf. Paul and Hall (2013), Figure 11.

And we similarly find that there is a genuine process from *B* to *C*, namely:

t_1: *B*

t_2: *C*

These two processes are illustrated below:

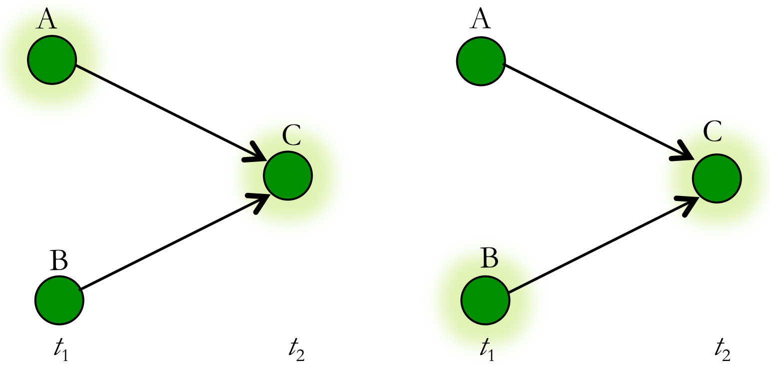


Figure 4*

Thus, we find – as desired – that both *A* and *B* satisfy the condition of process-connection.

3. Omission-involving causation

In this section, I show how the condition of process-connection is able to deal successfully with omission-involving causation. I begin by considering the three standard kinds of omission-involving causation: causation by omission (section 3.1), prevention (section 3.2), and double prevention (section 3.3). Next, I consider corresponding cases of *redundant* omission-involving causation: redundant causation by omission (section 3.4), redundant prevention (section 3.5), and redundant double prevention (section 3.6). Finally, I revisit Hall's case of preemptive prevention (section 3.7).

It has sometimes been suggested that causation involving omissions should be given a separate treatment from causation involving only positive events. For example, this suggestion is part of Hall's two concepts account of causation.[11] Cases of redundant omission-involving causation provide an important argument against this approach: they show that the cases of redundant causation that present challenges for an account of causation involving only positive events have analogues involving omissions. Once we recognise this, it is sensible to look for an account that can give a unified treatment of redundant causation, whether it involves omissions or not. As I hope to show, the condition of process-connection allows us to do just that.

3.1 Causation by omission

As a first illustration of how the condition of process-connection applies to omission-involving causation, consider the following simple case of causation by omission:[12]

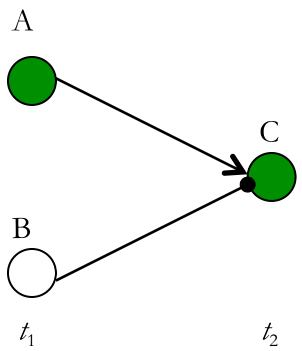


Figure 5

It is intuitively clear that C has two causes: A and $\neg B$. Indeed, C depends counterfactually on A and on $\neg B$: if **A** had not fired, **C** would not have fired; and if **B** had not *failed* to fire, **C** would not have fired.[13]

Correspondingly, we find that A and $\neg B$ both satisfy the condition of process-connection. To see this, note that based on the time-series T: $t_1 < t_2$, there is a genuine process from A to C, namely:

---

[11] See Hall (2004*b*).

[12] Cf. Paul and Hall (2013), Figure 3.

[13] Cf. Paul and Hall (2013), p. 174.

t_1: A (together with $\neg B$)

t_2: C

And there is a genuine process from $\neg B$ to C, namely:

t_1: $\neg B$ (together with A)

t_2: C

Thus, A and $\neg B$ both satisfy the condition of process-connection. In this way, the condition easily yields the intuitively correct result.

3.2 Prevention

Next, let us consider a case of causation *of* omission, i.e. prevention. Such a case is illustrated in the neuron diagram below:[14]

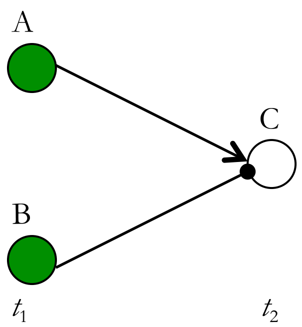


Figure 6

Intuitively, B is a cause of $\neg C$. Indeed, **C**'s failure to fire depends counterfactually on **B**'s firing: if **B** had not fired, **C** would not have failed to fire.[15]

Correspondingly, we find that there is a genuine process from B to $\neg C$ based on the time-series T: $t_1 < t_2$, namely:

[14] Cf. Paul and Hall (2013), Figure 22.

[15] Paul and Hall (2013), p. 174.

t_1: B

t_2: $\neg C$

Thus we find – as we should – that B is process-connected to $\neg C$.

3.3 Double prevention

The third standard kind of omission-involving causation is double prevention. The neuron diagram below illustrates the structure of such a case:[16]

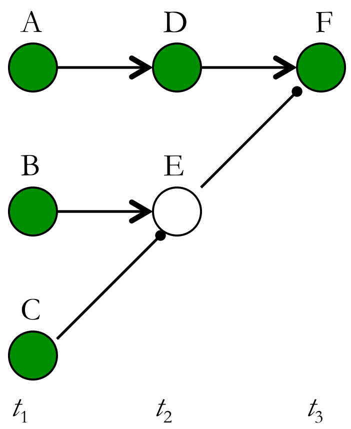


Figure 7

Intuitively C, together with A, is a cause of F.[17] Indeed, F depends counterfactually on C: if C had not occurred, F would not have occurred. The case earns its title as a case of omission-involving causation because an omission, namely $\neg E$, acts as an intermediary between C and F: embedded in the figure, we find a standard case of prevention, where C causes $\neg E$ (cf. Figure 6 above), and a standard case of causation by omission, where $\neg E$ is a cause of F (cf. Figure 5 above).[18]

Correspondingly, we find that there is a genuine process from C to F based on the time-series T: $t_1 < t_2 < t_3$, namely:

[16] Cf. Paul and Hall (2013), Figure 29.

[17] Paul and Hall (2013), p. 175.

[18] Schaffer (2000*a*) gives several examples of real life cases with exactly this structure.

t_1: C

t_2: $\neg E$ (together with D)

t_3: F

This process is illustrated below:

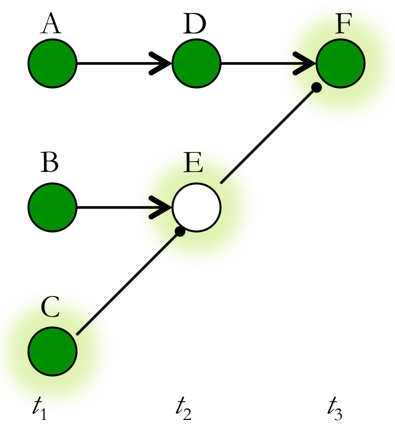


Figure 7*

We thus get the intuitively correct result that C is process-connected to F.

3.4 Redundant causation by omission

I now turn to more complex cases that combine redundant causation with omission-involving causation. My first example is a case of redundant causation by omission:[19]

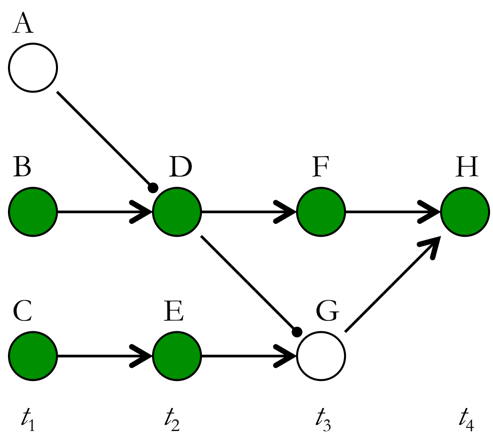


Figure 8

[19] Cf. Paul and Hall (2013), Figure 38.

Embedded within the figure, there is a standard case of causation by omission: ¬*A* is a cause of *D*. And there is also, embedded within the figure, a standard case of early preemption: *D* is a preempting cause of *H*, whereas *E* is a preempted backup. Based on this, it seems intuitively correct to say that ¬*A* is a preempting cause of *H*, and that *C* is merely a preempted backup.[20]

The condition of process-connection easily accommodates these intuitive judgements. To see this, note that there is a genuine process from ¬*A* to *H* based on the time-series *T*: $t_1 < t_2 < t_3 < t_4$, namely:

t_1:	¬*A*	(together with *B*)
t_2:	*D*	
t_3:	*F*	
t_4:	*H*	

This process is illustrated below:

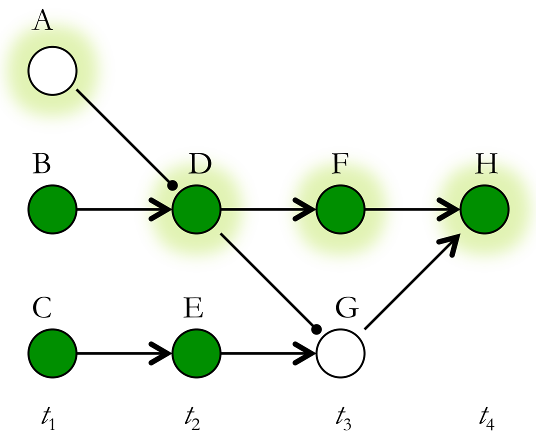


Figure 8*

Thus, we find that ¬*A* is process-connected to *H*. Furthermore, we find that *C* is *not* process-connected to *H*, since there is no genuine process from *C* to *H*.

[20] Cf. Paul and Hall (2013), pp. 187 and 214.

3.5 Redundant prevention

As our next example, consider the case of redundant prevention illustrated below:[21]

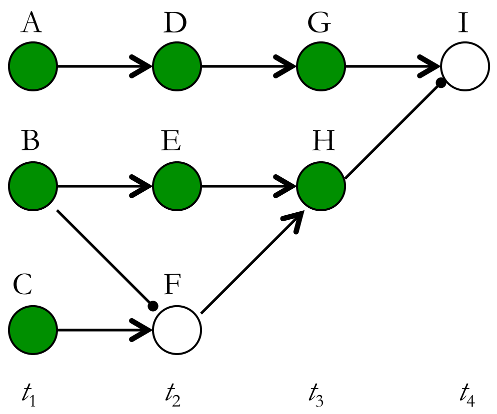


Figure 9

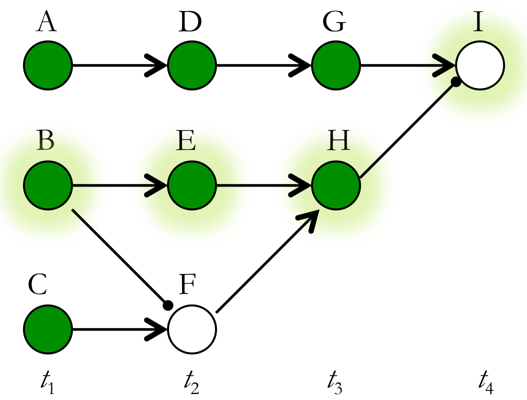
To see that this is indeed a case of redundant prevention, note that there is, embedded within the figure, a standard case of early preemption: *B* is a preempting cause of *H*, while *C* is a preempted backup. And note that there is also, embedded within the figure, a standard case of prevention where *H* is a cause of ¬*I*. Intuitively, it therefore seems right to say that *B* is a preempting cause of ¬*I*, whereas *C* is merely a preempted backup.[22]

This fits exactly with the results we get when we apply the condition of process-connection. To see this, note that there is a genuine process from *B* to ¬*I*, based on the time-series *T*: $t_1 < t_2 < t_3 < t_4$, namely:

t_1: *B*
t_2: *E*
t_3: *H*
t_4: ¬*I*

[21] Cf. Paul and Hall (2013), Figure 43.
[22] Cf. Paul and Hall (2013), pp. 213-14.

This process is illustrated below:



F **Figure 9***

We thus find that *B* is process-connected to $\neg I$. Furthermore, we find that *C* is not process-connected to $\neg I$, since there is no genuine process from *C* to $\neg I$.

3.6 Redundant double prevention

Next, consider the case of redundant double prevention illustrated below:[23]

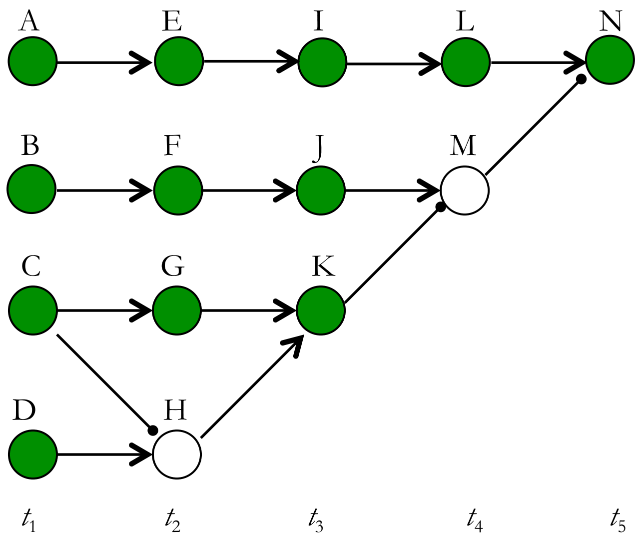


Figure 10

[23] Cf. Paul and Hall (2013), Figure 44.

This figure combines a case of early preemption with a case of double prevention: *C* is a preempting cause of *K*, while *D* is a preempted backup. And *K* in turn causes *N* via double prevention. On this basis, it seems intuitively clear that *C* is a preempting cause of *N*, whereas *D* is a preempted backup.[24]

This fits exactly with what we find when we apply the condition of process-connection: we here find that there is a genuine process from *C* to *N*, namely:

t_1:	C	
t_2:	G	
t_3:	K	
t_4:	$\neg M$	(together with *L*)
t_5:	N	

This process is illustrated below:

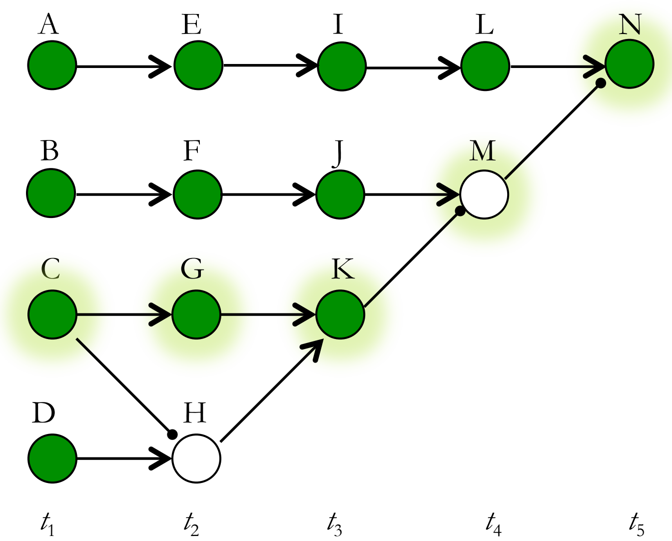


Figure 10*

[24] Cf. Paul and Hall (2013), p. 220.

Thus, we find that *C* is process-connected to *N*. By contrast, there is no genuine process from *D* to *N*, and so *D* does not satisfy the condition of process-connection.

3.7 Hall's case

As a final example of how the condition of process-connection allows us to handle cases of omission-involving redundant causation, let us revisit the case of redundant omission-involving causation discussed in Chapter 2 section 2.2.2. For ease of reference, I repeat the figure here:[25]

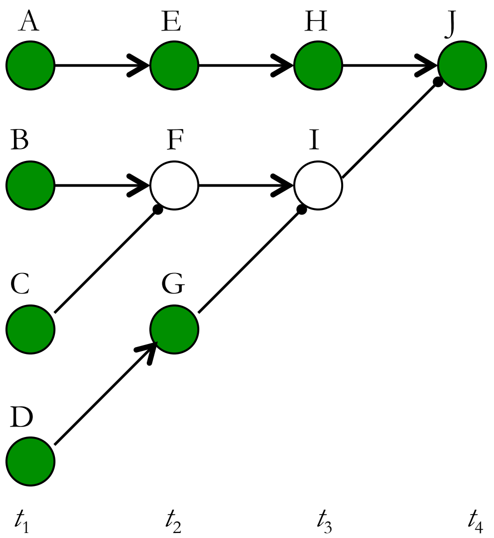


Figure 11

As we have seen, the intuitively correct verdict on this case is that *C* is a cause of *J*. As Hall writes:

> '[*D*] notwithstanding, it is *C* that *in fact* cancels the threat to [*J*], and canceling a threat is one way to be a cause.'[26]

[25] Figure from Hall (2007*b*), p. 52.

[26] Hall (2007*b*), p. 52.

The condition of process-connection easily accommodates this verdict: there is a genuine process from *C* to *J*, namely:

t_1:	C	
t_2:	$\neg F$	
t_3:	$\neg I$	(together with H)
t_4:	J	

This process is illustrated below:

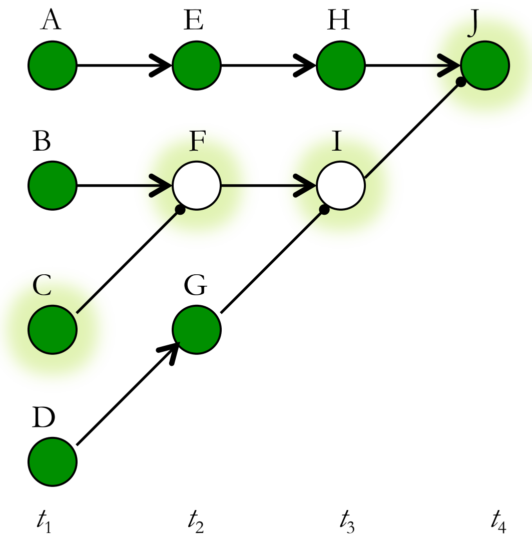


Figure 11*

Hence, we find – exactly as we should – that *C* is process-connected to *J*.

However, one might suggest that the present case puts pressure on my claim that the condition of process-connection allows us to distinguish preempting causes from preempted backups: one might well judge that *C* is a preempting cause of *J*, while *D* is merely a preempted backup. As I have presented the case, however, there is also a genuine process from *D* to *J*, namely:

t_1:	D
t_2:	G

t_3: $\neg I$ (together with H)

t_4: J

This process is illustrated below:

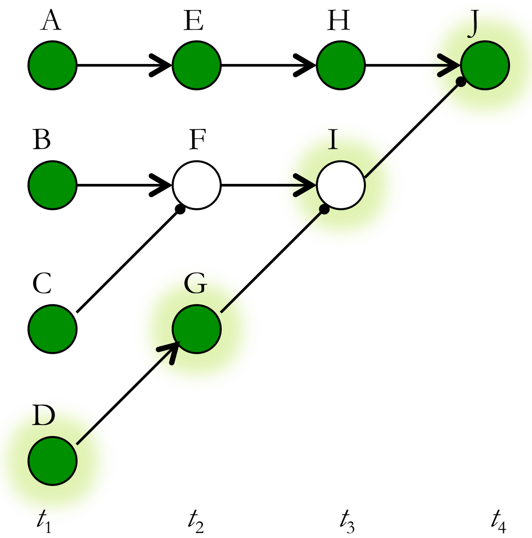


Figure 11**

Thus, we find that D is also process-connected to J.

My response has two parts. First, I believe that as it is presented here, this is a case of *overdetermination* rather than preemption: $\neg I$ is in fact *overdetermined* by $\neg F$ and G, which are caused, respectively, by C and D.

Second, I suggest that the intuition that D is merely a preempted backup is based on an intuitive distinction between two different ways in which a neuron can fail to fire: it can fail to fire because it receives no stimulatory signals, or it can fail to fire because a stimulatory signal is blocked by an inhibitory signal. This distinction is not in fact present in the figure above. However, we can easily adapt the case so that it reflects this distinction, by making the following additions to our usual conventions for neuron diagrams: let a neuron be blank just in case it receives no stimulatory signals, and let it fire in waves just in case it receives at least one stimulatory signal together with an inhibitory signal. Furthermore, add to the neuron laws that a neuron that fires in waves sends no outgoing signals. With these additions, we now get the following version of the case:

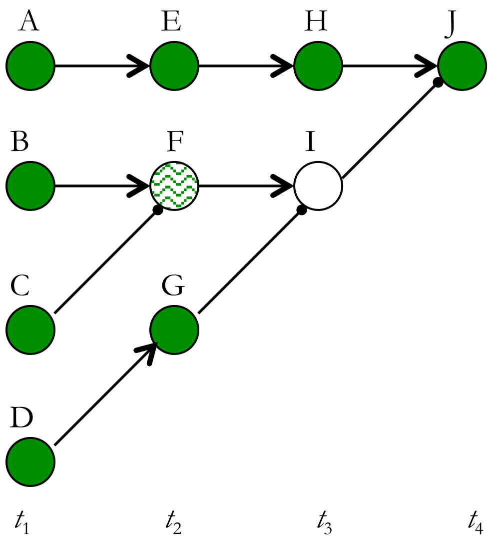


Figure 12

This is indeed a case of preemption, where *C* is a preempting cause of *J*, while *D* is a preempted backup. Correspondingly, we find that there is a genuine process from *C* to *J*, namely:

t_1:	*C*	(together with *B*)
t_2:	*F-waves*	
t_3:	$\neg I$	(together with *H*)
t_4:	*J*	

This process is illustrated below:

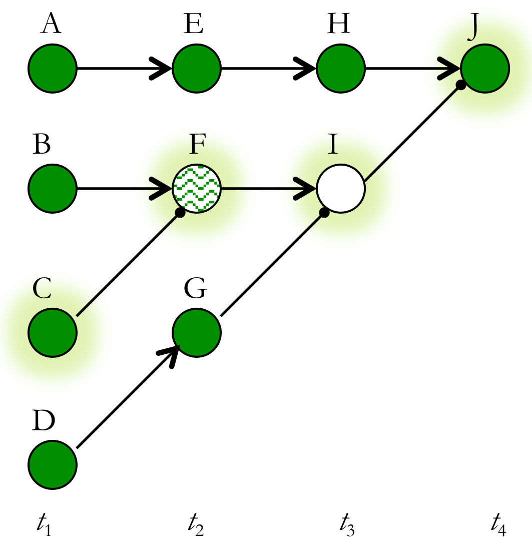


Figure 12*

Thus, *C* is process-connected to *J*. But in this case, there is *no* genuine process from *D* to *J*, and so we find – as we should – that *D* is *not* process-connected to *J*. Thus, the condition of process-connection successfully distinguishes the preempting cause from the preempted backup.

4. Commensuration between cause and effect

As a final application of the condition of process-connection, I will consider a group of cases that demonstrate the need to ensure commensuration between cause and effect. I begin by considering two counterexamples to the sufficiency of counterfactual dependence for causation (section 4.1 and 4.2). And finally, I consider an apparent counterexample to the transitivity of causation (section 4.3).

4.1 Scarlet

Let us begin by considering our first counterexample to the sufficiency of counterfactual dependence for causation – namely *Scarlet* (see Chapter 2 section 1.3.1). For ease of reference, I repeat the case here:

> *Scarlet*: The pigeon Sophia has been conditioned to peck at scarlet to the exclusion of all other colours. She is presented with a scarlet triangle and pecks at it.[27]

The neuron diagram below represents the structure of this case:

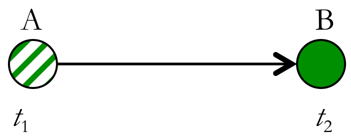


Figure 13

Neuron **A** can here fire in two different ways: in stripes or in uniform green. As shown in the figure, it in fact fires in stripes. Let *A* be the event that is essentially **A**'s firing (but where it is not essential whether **A** fires in stripes or uniform green), and let *A-stripes* be the event that is essentially **A**'s firing in stripes. Furthermore, suppose that the neuron laws are such that **B** fires if and only if **A** fires in stripes.

The structure of this case corresponds perfectly to the structure of *Scarlet*: *A* represents the triangle's being red, *A-stripes* represents the triangle's being scarlet, and *B* represents Sophia's pecking.

As we have seen, this case presents a counterexample to the sufficiency of counterfactual dependence for causation: *B* (Sophia's pecking) depends counterfactually on *A-stripes* (the triangle's being scarlet). Correspondingly, we judge that *A-stripes* is a cause of *B*. So far, no counterexample. But *B* also depends counterfactually on *A* (the triangle's being red). However, we are reluctant to say that *A* is a cause of *B*.

The condition of process-connection accommodates these intuitive verdicts – and thereby goes against the principle that counterfactual dependence is sufficient for causation: *A-stripes* is process-connected to *B*,

[27] This case is closely based on a case presented in Yablo (1992*a*), p. 257. For similar cases, see Yablo (1992*b*), p. 415, and Sartorio (2010), pp. 266-69.

whereas A is not. To see this, note that there is a genuine process from *A-stripes* to B, namely:

t_1: *A-stripes*
t_2: B

By contrast, there is no genuine process from A to B: the only minimally sufficient set for B is {*A-stripes*}, and A does not belong to *A-stripes*.
In this way, the condition of process-connection respects the intuitive verdict that *A-stripes* is a cause of B, while A is not.

4.2 Red

The condition of process-connection similarly respects our intuitive verdict on the second counterexample to the sufficiency of counterfactual dependence (see Chapter 2 section 1.3.1). This example goes as follows:

> *Red*: The pigeon Delia has been conditioned to peck at red to the exclusion of all other colours. She is presented with a scarlet triangle and pecks at it. In the lab where she is, the researchers use just two colours – scarlet and emerald. If Delia had not been presented with a scarlet triangle, she would have been presented with an emerald triangle.[28]

We may use the same neuron diagram as above to represent the structure of this case:

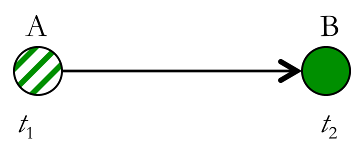


Figure 13

[28] This case is closely based on a case presented in Yablo (1992*a*), p. 257. For similar cases, see Yablo (1992*b*), p. 417, and Sartorio (2010), pp. 264-66.

The difference between this case and the previous is that the neuron laws are different: in the present case, the laws are such that **B** fires just in case **A** fires (whether **A** fires in stripes or in uniform green). Furthermore, in the closest world(s) where *A-stripes* does not occur, *A* does not occur either. Thus, the neuron diagram below represents what would have happened if *A-stripes* had not occurred:

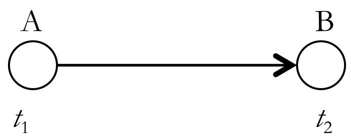


Figure 13*

This case presents another counterexample to the sufficiency of counterfactual dependence for causation: *B* (Delia's pecking) depends counterfactually on *A-stripes* (the triangle's being scarlet). Intuitively, however, we are reluctant to say that *A-stripes* is a cause of *B* – intuitively, the cause is simply *A* (the triangle's being *red*); it does not matter that **A** fires in stripes (that the triangle is *scarlet*).

The condition of process-connection respects this intuitive verdict: *A* is process-connected to *B*, whereas *A-stripes* is not. To see this, note that there is exactly one minimally sufficient set for *B*, namely $\{A\}$. It follows from this that there is a genuine process from *A* to *B*, namely:

t_1: *A*
t_2: *B*

And it also follows that there is no genuine process from *A-stripes* to *B*. Thus, we find that *A* is process-connected to *B*, whereas *A-stripes* is not.

4.3 Apparent counterexample to the transitivity of causation

Finally, let us now turn to the apparent counterexample to the transitivity of causation discussed in Chapter 4 section 6. For ease of reference, I repeat the case here:

> *Skiing accident*: while skiing, Suzy breaks her right wrist. The next day, she writes a philosophy paper, which is subsequently accepted for publication. Since Suzy's right wrist is broken, she writes the paper by typing with her left hand. And as she is not used to writing this way, she develops a cramp in her left hand.[29]

The structure of this case may be represented as follows:[30]

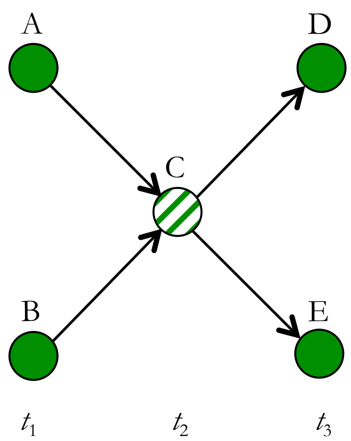


Figure 14

A here represents Suzy's having the idea for her paper, *B* represents her breaking her right wrist in the skiing accident, *C* represents her writing the paper, *D* represents the paper's being accepted for publication and *E* represents Suzy getting a cramp in her left hand. Note that **C** can fire in two different ways: in uniform green and in stripes. **C**'s firing in uniform green represents Suzy's writing her paper as she normally would – by typing with both hands; **C**'s firing in stripes represents Suzy's writing the paper by typing only with her left hand.

The neuron laws are such that **C** fires in uniform green if and only if it receives a stimulatory signal from **A** and no stimulatory signal from **B**; and **C**

[29] This case is closely based on a case presented in Paul (2004*a*). For discussion of similar cases, see also Ehring (2009), pp. 403-4; McDonnell (forthcoming); Paul and Hall (2013), pp. 237-44; Schaffer (2016), pp. 13-14, and Woodward (1984), pp. 234-46.

[30] Cf. Paul and Hall (2013), Figure 48.

fires in stripes if and only if it receives stimulatory signals from both **A** and **B**. Furthermore, **D** fires if and only if **C** fires. By contrast, **E** is sensitive to the way in which **C** fires – and so, **E** fires if and only if **C** fires in stripes.

Finally, let C be the event that is essentially **C**'s firing (Suzy's writing the paper), but only accidentally **C**'s firing in stripes; and let *C-stripes* be the event that is essentially **C**'s firing in stripes (Suzy's writing the paper by typing with her left hand).

As we have seen in Chapter 4 section 6, this case presents an apparent counterexample to the transitivity of causation. As we have also seen, the first step in resisting this counterexample is to adopt a conception of the causal relata that allows us to hold that C (Suzy's writing the paper) is *not* identical to *C-stripes* (Suzy's writing the paper by typing with her left hand). This is exactly what my proposed conception of the causal relata allows us to do.

To resist the counterexample, we also need an account of causation that delivers the following verdicts: B is *not* a cause of C, though it is a cause of the more fragile *C-stripes*. And in turn C is a cause of D, while *C-stripes* is not.[31] The condition of process-connection delivers exactly these results:

B does not belong to a minimally sufficient set for C, since the only minimally sufficient set for C is $\{A\}$. Thus, B is *not* process-connected to C. By contrast, B is process-connected to *C-stripes* via the following process:

t_1: B (together with A)
t_2: *C-stripes*

Furthermore, we find that C is process-connected to D via the following process:

t_2: C
t_3: D

[31] For discussion, see McDonnell (forthcoming).

By contrast, *C-stripes* is *not* process-connected to *D*: though sufficient, the set {*C-stripes*} is not minimally sufficient for *D*, since it is a strictly more fragile version of {*C*}, which is already sufficient. Thus, *C-stripes* does not belong to a minimally sufficient set for *D*.

Given these results, transitivity simply does not apply. In this way, the condition of process-connection allows us to resist the apparent counterexample.

5. Conclusion

In this chapter, we have seen the main applications of the condition of process-connection. First, we have seen how the condition enables us to distinguish mere correlation from causation. Second, we have seen how it enables us to distinguish genuine causes from preempted backups in the four main kinds of redundant causation – early preemption, late preemption, trumping preemption, and symmetric overdetermination: in all of these cases, the condition of process-connection is satisfied by the genuine cause(s), but fails to be satisfied by their preempted backups.

Third, we have seen how process-connection can effortlessly accommodate omission-involving causation, delivering intuitively correct results in cases of causation by omission, prevention, and double prevention, as well as in cases of omission-involving redundant causation. And finally, we have seen how the condition of process-connection delivers intuitively correct results in cases where we need commensuration between cause and effect.

Based on the cases considered so far, one might indeed be tempted to suggest that process-connection is both necessary and sufficient for causation. That, however, would be a mistake: there are clear cases where an event *c* is process-connected to a later event *e*, but where we intuitively judge that *c* is *not* a cause of *e*. Indeed, the counterexamples to the transitivity and intrinsicness of causation considered in Chapter 2 section 1.3.2 and 1.3.3 are cases of just this kind. To accommodate our intuitive judgements in such cases, we need to supplement the necessary condition of process-connection with a further necessary condition – namely, the condition of *security-dependence*.

PART IV

Security-dependence

8

Defining security-dependence

A cause must somehow make a difference to its effect. Indeed, this was the guiding idea behind Lewis's counterfactual theory of causation:

> 'We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it.'[1]

In this chapter, I set out my second necessary condition for causation – the condition of *security-dependence* – which is intended to capture this idea.

The condition of security-dependence is a counterfactual-based condition. Thus, the first step in setting out this condition is to give a recipe for evaluating the relevant counterfactuals – what we may call *causal counterfactuals*. These counterfactuals have antecedents of the form 'if c had not occurred, then . . .', or, more generally, 'if c_1, c_2, . . . , c_n had not occurred, then . . .', where c_1, c_2, . . . , c_n are contemporaneous events. And, importantly, they require a non-backtracking reading.

Maudlin, Hall, and Paul have developed a recipe for evaluating counterfactuals that is specifically tailored to causal counterfactuals.[2] The recipe for evaluating causal counterfactuals that I present here is closely based on their proposal. However, I suggest an important modification: it is usually assumed that counterfactuals should be evaluated relative to the entire space of possible worlds. By contrast, I suggest that causal counterfactuals should be evaluated relative to a *restricted* space of nomologically possible worlds –

[1] Lewis (1986*d*), pp. 160-62.
[2] Hall (2007*b*), pp. 18-21; Maudlin (2007*b*); Paul and Hall (2013), pp. 43-53.

namely, a contextually determined *possibility horizon* of the kind presented in Chapter 5.

Based on my proposed recipe for evaluating causal counterfactuals within a possibility horizon, I next set out two relations that I suggest are both crucial for understanding causation: the ternary relation of counterfactual dependence within a possibility horizon, and the ternary relation of security-dependence within a possibility horizon.

In paradigm cases of difference-making, the effect depends counterfactually on its cause within the contextually relevant possibility horizon $\mathscr{H}$: in such cases, it is true within $\mathscr{H}$ that the effect simply would not have occurred, if the cause had not occurred. In cases of redundant causation, on the other hand, the effect would still have occurred – sometimes at exactly the same time and in exactly the same manner – even if the cause had not occurred. Such cases demonstrate that counterfactual dependence within a possibility horizon is not a necessary condition for causation.

To find a necessary condition for causation that captures the idea that a cause is something that makes a difference, we therefore need to find a difference-making relation that is entailed by counterfactual dependence, but which also holds in just the right selection of further cases. To do this, my suggestion is that we need to pay attention to the *modal* features of the effect: we need to pay attention to *how easily the effect could have failed to happen*.

My proposal is that a cause must make a difference to how easily its effect could have failed to happen. More precisely, it must make the right kind of difference: it must be the case that if the cause had not occurred, the effect could *more easily* have failed to happen.[3] When this relationship holds between two events within a possibility horizon $\mathscr{H}$ – roughly, when it is true within $\mathscr{H}$ that *e* could more easily have failed to happen if an earlier event *c* had not occurred – I will say that *e security-depends on c within possibility horizon $\mathscr{H}$.*

[3] For different ways to develop this idea, see Sartorio (2005) and (2006), and Yablo (2004). See also Woodward (2006).

As we shall see, counterfactual dependence within a possibility horizon entails security-dependence within that same possibility horizon. Thus, the condition of security-dependence is automatically satisfied in all our paradigm cases of difference-making, where the effect depends counterfactually on its cause(s) within the counterfactually determined possibility horizon. In addition, we shall see in Chapter 9 and 10 that the condition of security-dependence is satisfied in just the right selection of further cases: it is satisfied in cases of redundant causation, but not in cases where transitivity and intrinsicness fail, etc.

In the following, I proceed in four steps. First, I set out my general recipe for evaluating causal counterfactuals (section 1). Building on this, I define the relations of counterfactual dependence within a possibility horizon (section 2), and security-dependence within a possibility horizon (section 3).

1. Evaluating causal counterfactuals within a possibility horizon

In this section, I present my recipe for evaluating causal counterfactuals within a possibility horizon. This recipe is closely based on the recipe developed by Maudlin, Hall, and Paul.[4] I set out my general recipe in section 1.1. I then illustrate two features of this recipe: that it ensures that our causal counterfactuals receive a non-backtracking reading (section 1.2), and that it resolves what we may call the problem of excision (section 1.3).

1.1 The recipe

The Maudlin-Hall-Paul recipe for evaluating causal counterfactuals makes a crucial departure from e.g. Lewis's semantics for counterfactuals: Lewis's semantics is concerned with similarity between entire worlds.[5] By contrast, the Maudlin-Hall-Paul proposal focuses on similarity between states of worlds at times:

---

[4] Hall (2007*b*), pp. 18-21; Maudlin (2007*b*); Paul and Hall (2013), pp. 43-53.

[5] See e.g. Lewis (1973) and (1986*c*).

> 'Similarity enters in not as a relation between entire worlds, but as a relation between states of worlds at times.'[6]

This innovation has important advantages. In particular, it ensures that our causal counterfactuals receive a non-backtracking reading.[7] I therefore follow Maudlin, Hall, and Paul in focusing on relations of similarity between complete states of worlds at times. We may think of this in terms of *closeness-at-a-time*: two worlds w_i and w_j are close-at-time-t to the extent that the complete state of w_i at t is similar to the complete state of w_j at t.

There is, however, another respect in which the Maudlin-Hall-Paul recipe does follow Lewis: by assuming that causal counterfactuals should be evaluated relative to the entire space of possible worlds. By contrast, I suggest that causal counterfactuals should be evaluated within a contextually determined *possibility horizon* (see Chapter 5). Thus, I propose the following definition:[8]

> *Causal counterfactuals*: let $c_1, c_2, \ldots, c_n$ be instantaneous events occurring at time t. Then the causal counterfactual 'if none of the events $c_1, c_2, \ldots, c_n$ had occurred, then it would have been the case that p' is true within possibility horizon $\mathscr{H}$ iff
>
> a) there is at least one world in $\mathscr{H}$ where none of $c_1, c_2, \ldots, c_n$ occur, and
>
> b) in the closest-at-t world(s) in $\mathscr{H}$ where none of $c_1, c_2, \ldots, c_n$ occur, it is the case that p.

An important feature of the above definition is the way in which it treats causal counterfactuals with antecedents that are *not* satisfied in any world in the

[6] Paul and Hall (2013), p. 48.

[7] See e.g. Paul and Hall (2013), pp. 48-49.

[8] Note that this and the following definitions are made from the standpoint of the actual world. However, any world w can play the role of our actual world, and so the definitions are fully general, and can be used to evaluate causal counterfactuals from the standpoint of any world w. Cf. Lewis (1986*d*), p. 163.

relevant possibility horizon $\mathscr{H}$. To appreciate this point, it will be useful to compare my proposal with Lewis's:

On Lewis's semantics for counterfactuals, a counterfactual with a necessarily false antecedent counts as vacuously true.[9] However, Lewis ensures that such vacuously true counterfactuals never play a role in his account of causation: he builds into his account of events that events occur contingently, so that for any event *c*, there is a possible world where *c* does not occur – and this ensures that counterfactuals of the form 'if *c* had not occurred, . . .' never have necessarily false antecedents.[10]

By contrast, my notion of a possibility horizon allows for cases where an event *c* occurs in *every* world within the contextually relevant possibility horizon. Indeed, my proposal on how context selects a possibility horizon ensures that whenever an event *c* is categorised as a default event, *c* occurs in every world within the contextually relevant possibility horizon. In such cases, the antecedent of the causal counterfactual 'if *c* had not occurred, . . .' is not satisfied in any world within the relevant possibility horizon $\mathscr{H}$. My suggestion is that we should count such a counterfactual as *not* being true on the reading that is required for an account of causation, for the reason that it is concerned with a possibility that is deemed *irrelevant* within the contextually determined possibility horizon. This suggestion departs from Lewis's treatment of counterfactuals with necessarily false antecedents. However, it keeps what is important, namely the idea that only non-vacuously true counterfactuals can support causal claims.

To make the above recipe for evaluating counterfactuals more precise, I now need to say more about distance-at-a-time. I adopt the following definition of what it takes for a world to be among the closest-at-*t* worlds where none of the events c_1, c_2, . . . , c_n occur, leaving open the possibility that

---

[9] Lewis (1973), pp. 24-26.

[10] Lewis (1986*g*), p. 243.

there may be more than one closest-at-t world where none of these events occur:[11]

> *Closest-at-t*: from the standpoint of @, a world w is among the closest-at-t worlds in $\mathscr{H}$ where none of the events $c_1, c_2, \ldots, c_n$ occur iff
>
> a) w belongs to $\mathscr{H}$ and none of $c_1, c_2, \ldots, c_n$ occur in w, and
>
> b) there is no world w^* such that
>
>   i) w^* belongs to $\mathscr{H}$ and none of $c_1, c_2, \ldots, c_n$ occur in w^*, and
>
>   ii) the distance-at-t between @ and w^* is *strictly shorter* than the distance-at-t between @ and w^*.

The distance-at-a-time between worlds is a matter of overall similarity between complete states. To some degree, I will in the following rely on our intuitive ability to make the required judgements of overall similarity. Furthermore, I suggest two principles that we may use as rules of thumb.[12] To set out these principles, it will be useful to have the following notion of a *difference*:

> *Difference*: let c be an event that occurs at t in w_i, but does not occur at t in w_j. Then c is a *difference* from the state-at-t of w_i to the state-at-t of w_j.

The first rule of thumb is encapsulated in the following principle:

> *Principle I*: if the set of differences from the state-at-t of w_i to the state-at-t of w_j is *the same as* the set of differences from the state-at-t of w_k to the state-at-t of w_l, then the distance-at-t from w_i to w_j = the distance-at-t from w_k to w_l.

---

[11] The proposed definition does not allow for cases where there is an infinite series of closer and closer worlds, but no closest world(s). With a bit of added complexity, however, the definition can easily be modified to accommodate such cases. For discussion, see Lewis (1973), pp. 19-21, and Collins, Hall, and Paul (2004*b*), p. 4.

[12] Both of these principles require us to quantify over events. If we do not include enough events within our domain of quantification, the principles yield intuitively false results. If we include too many events, we rarely if ever find circumstances where the principles can be applied. Given our usual restrictions on the domain of instantaneous events, however, the principles are applicable in many cases, and yield intuitively correct results.

To illustrate how this principle applies, consider the neuron diagram below:

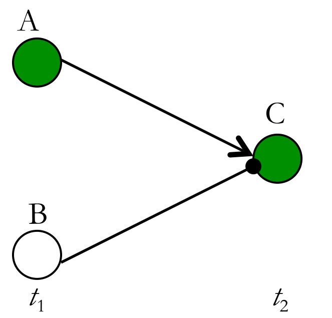


Figure 1

And consider possibility horizon $\mathscr{H}$, with the following four possible worlds:

	Complete state at time t_1:	*Events at t_1*:		t_2:
@:	**A** fires, **B** does not fire	A,	$\neg B$,	C
w_1:	**A** does not fire, **B** does not fire	A,	$\neg B$,	C
w_2:	**A** fires, **B** fires	A,	$\neg B$,	C
w_3:	**A** does not fire, **B** fires	A,	$\neg B$,	C

Let us, for example, compare the distance-at-t_1 from @ to w_1 with the distance-at-t_1 from w_2 to w_3. In keeping with our usual practice when working with neuron diagrams (cf. Chapter 4 section 2.3), we restrict our quantifiers to maximally temporally fragile events based on the firing or failure to fire of a single neuron. Based on this restriction, we find that the set of differences from the state-at-t_1 of @ to the state-at-t_1 of w_1 is $\{A\}$: A is the only event that occurs at t_1 in @, and does not occur at t_1 in w_1. The set of differences from the state-at-t_1 of w_2 to the state-at-t_1 of w_3 is exactly the same, namely $\{A\}$. By *Principle I*, it therefore follows that the distance-at-t_1 from @ to w_1 is exactly the same as the distance-at-t_1 from w_2 to w_3.

The second rule of thumb is encapsulated by the following principle:

> *Principle II*: if the set of differences from the state-at-t of w_i to the state-at-t of w_j is a *proper subset* of the set of differences from the state-at-t of w_k to the state-at-t of w_l, then the distance-at-t from w_i to w_j < the distance-at-t from w_k to w_l.

As an illustration of how this principle applies, let us again consider the four worlds listed above. In particular, let us now compare the distance-at-t_1 from @ to w_1 with the distance-at-t_1 from @ to w_3. We have already seen that the set of differences from the state-at-t_1 of @ to the state-at-t_1 of w_1 is $\{A\}$. Furthermore, the set of differences from the state-at-t_1 of @ to the state-at-t_1 of w_3 is $\{A, \neg B\}$: A and $\neg B$ are the only events that occur in @ at t_1, and do not occur in w_3 at t_1. Clearly, $\{A\}$ is a proper subset of $\{A, \neg B\}$. Thus, it follows from *Principle II* that the distance-at-t_1 from @ to w_1 is strictly shorter than the distance-at-t_1 from @ to w_3.

Together, the two principles give a great deal of information about comparative distance-at-a-time. This is illustrated in the figure below, which sums up all applications of the two principles to the four worlds in $\mathscr{H}$.

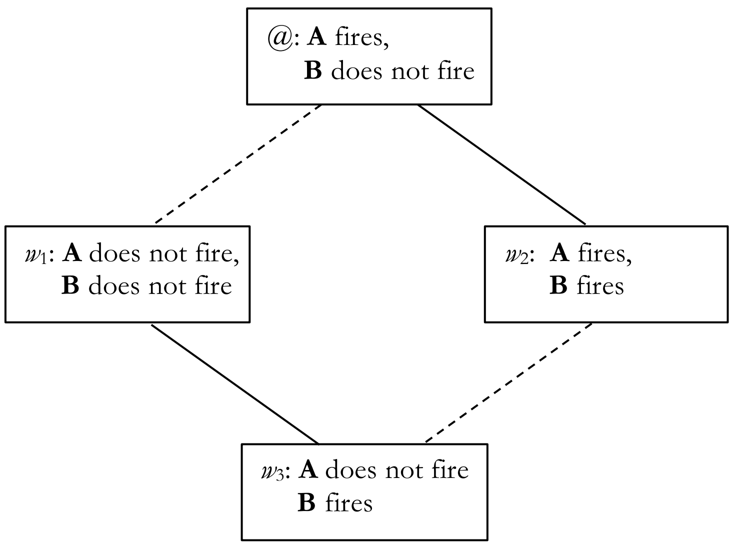


Figure 2

When two worlds w_i and w_j in this figure are connected by a line with the same typography as the line connecting two worlds w_k and w_l, this means that the distance-at-t_1 between w_i and w_j is exactly the same as the distance-at-t_1 between w_k and w_l. And when the shortest route from a world w_i to a world w_k goes via a world w_j, this means that the distance-at-t_1 from w_i to w_j is strictly shorter than the distance-at-t_1 from w_i to w_k.

Although the two principles help us settle many questions about comparative distance-at-a-time, however, they do not settle all such questions: for example, they leave it unresolved how the distance-at-t_1 from @ to w_1 compares to the distance-at-t_1 from @ to w_2. Such questions are determined directly by the relevant relation of overall similarity, and in making judgements about such cases, we must rely on our intuitive understanding of this relation of overall similarity.

In the following two sections, I will now illustrate two advantages of my proposed recipe for evaluating causal counterfactuals: that it ensures that our causal counterfactuals receive a non-backtracking reading, and that it resolves what we may call the problem of excision.

1.2 Ensuring a non-backtracking reading

It is an important advantage of the Maudlin-Hall-Paul recipe that it ensures that our causal counterfactuals receive a non-backtracking reading.[13] To appreciate the importance of this, consider our standard case of correlation without causation:

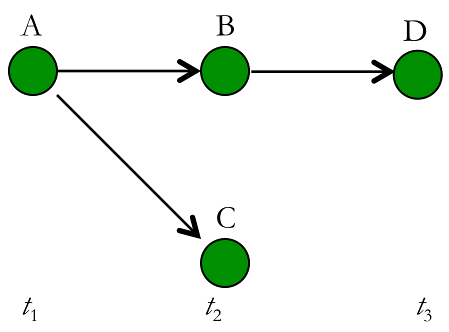


Figure 3

[13] See e.g. Paul and Hall (2013), pp. 48-49.

In particular, consider the counterfactual claim:

(P) If *C* had not occurred, *D* would not have occurred.

On a backtracking reading, we would reason as follows: in order for *C* not to occur, *it would have to have been the case* that *A* did not occur. So if *C* had not occurred, *D* would not have occurred. If we want to maintain that counterfactual dependence is sufficient (or near enough) for causation, then it is important to avoid such a backtracking reading: otherwise, we are left with the obviously unacceptable result that *C* is a cause of *D*.[14]

The Maudlin-Hall-Paul recipe avoids such a backtracking reading by focusing on similarity between states of worlds at times, rather than similarity between entire worlds.[15] By taking over this innovation, my proposed recipe retains this advantage of the Maudlin-Hall-Paul approach. To see this, let us consider how to evaluate (P) on my proposed account:

Taking as our starting point what happens at time t_2 when *C* occurs, we arrive at a possibility horizon that includes the following four worlds, characterised by their complete state at time t_2 (in the right-hand column I also indicate how they evolve forward with respect to the occurrence of *D*):

	Complete state at time t_2:	*Events at* t_2:	t_3:
@:	**B** fires, **C** fires	*B*, *C*	*D*
w_1:	**B** does not fire, **C** fires	*B*, *C*	*D*
w_2:	**B** fires, **C** does not fire	*B*, *C*	*D*
w_3:	**B** does not fire, **C** does not fire	*B*, *C*	*D*

Call this possibility horizon $\mathscr{H}$. By applying the two principles set out above, we arrive at the following facts about the comparative distance-at-t_2 between these four worlds:

[14] Lewis (1986*d*), pp. 170-71; Paul and Hall (2013), pp. 71-73; see also Woodward (2003), p. 14.
[15] Paul and Hall (2013), pp. 48-49.

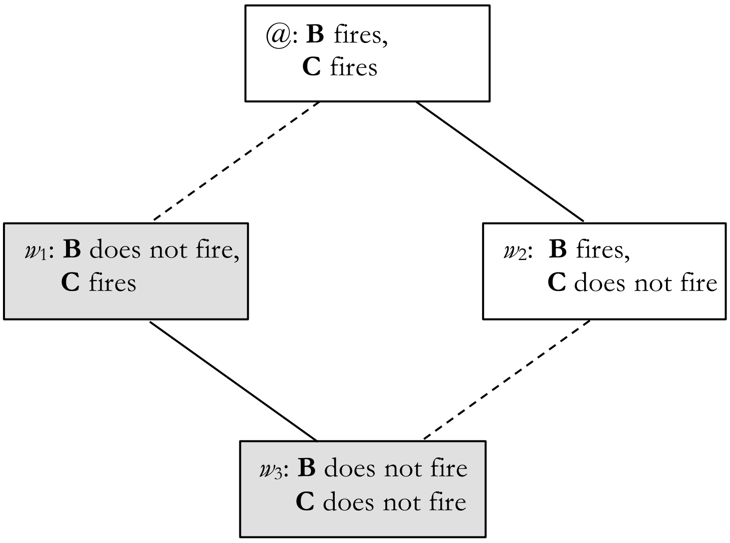


Figure 4

The worlds in which *D* occurs – namely, @ and w_2 – are here coloured white, while the worlds in which *D* fails to occur – namely, w_1 and w_3 – are coloured grey.

We now find that the closest-at-t_2 world in $\mathscr{H}$ where *C* does not occur is world w_2. *D* occurs in world w_2. Thus, we get the intuitively correct result, namely that (P) is false: if *C* had not occurred, *D* would still have occurred. In this way, my proposed recipe avoids a backtracking reading.

1.3 Resolving the problem of excision

A second advantage of my proposed recipe for evaluating causal counterfactuals is that it resolves what we may call the problem of excision. Lewis gives a nice statement of this problem as follows:

> 'What is the closest way to actuality for *C* not to occur? – It is for *C* to be replaced by a very similar event, one which is almost but not quite *C*, one that is just barely over the border that divides versions of *C* itself from its nearest alternatives. But if *C* itself is taken to be fairly fragile, then if *C* had not occurred and almost-*C* had occurred instead, very likely the effects of almost-*C* would have been much the same as the actual effects of *C*. So our causal

> counterfactual will not mean what we thought it meant, and it may well not have the truth value we thought it had.'[16]

Lewis then goes on to suggest the following solution:

> 'When asked to suppose counterfactually that C does not occur, we don't really look for the very closest possible world where C's conditions of occurrence are not quite satisfied. Rather, we imagine that C is completely and cleanly excised from history, leaving behind no fragment or approximation of itself.'[17]

Unfortunately, this suggestion is unsatisfactory for two reasons: first, it is not at all clear what it means that C should be 'completely and cleanly excised from history'.[18] Second, the solution seems *ad hoc*: why should we consider what happens in the worlds where C has been completely and cleanly excised from history (whatever exactly that means), when there are *closer* worlds where C does not occur?[19]

My proposal that causal counterfactuals should be evaluated within a contextually determined possibility horizon provides a simple solution to this problem: when an event c is categorised as a *deviant* event, this implies that extrapolation yields one or more alternatives c^*, c^{**}, . . . , to the occurrence of c. In that case, one of the restrictions that determines the contextually relevant possibility horizon $\mathscr{H}$ is that a world w belongs to $\mathscr{H}$ only if c or one of its alternatives c^*, c^{**}, . . . , occurs in w. This immediately yields the result that in the closest-at-t world(s) in $\mathscr{H}$ where c does not occur, c is replaced by one of its alternatives. When we evaluate causal counterfactuals within a possibility

---

[16] Lewis (2004*a*), p. 90.

[17] Lewis (2004*a*), p. 90.

[18] Cf. Hall (2007*b*), p. 20; Paul and Hall (2013), p. 51.

[19] This second problem also arises for the suggestion that we should consider the closest-at-t world(s) where the region where C occurs is returned to its default state (see Paul and Hall (2013), p. 51; Hall (2007*b*), p. 20): why should we consider these worlds when there are closer-at-t worlds where C does not occur?

horizon, then, the problem simply does not arise: within our contextually determined possibility horizon $\mathscr{H}$, the closest-at-*t* not-*c*-worlds just *are* worlds where *c* is replaced by one of its alternatives.

2. Counterfactual dependence

Based on our general recipe for evaluating causal counterfactuals within a possibility horizon, we may now capture the ternary relation of counterfactual dependence within a possibility horizon: *e* depends counterfactually on *c* within possibility horizon $\mathscr{H}$. I define this relation as follows:

> *Counterfactual dependence*: an event *e depends counterfactually* on an event *c* within possibility horizon $\mathscr{H}$ iff:
>
> a) *c* occurs at a time *t* strictly earlier than *e*,
>
> b) there is at least one not-*c*-world in $\mathscr{H}$, and
>
> c) in the closest-at-*t* not-*c*-world(s) in $\mathscr{H}$, *e* does not occur.

Note that condition b) and c) simply specify the conditions that have to be satisfied in order for the causal counterfactual 'if *c* had not occurred, then *e* would not have occurred' to be true within possibility horizon $\mathscr{H}$.

Definitions of counterfactual dependence standardly add a requirement that *c* and *e* must be distinct.[20] This is to exclude cases of logical rather than causal dependence, as when e^+ is a more fragile version of *e*, and the occurrence of e^+ therefore in some sense depends on the occurrence of *e*. In the above definition, I have left out this requirement because it is not needed: since *c* and *e* are instantaneous events, the requirement that *c* occurs strictly earlier than *e* automatically ensures that *c* and *e* are distinct.

In paradigm cases of difference-making, we find that the effect depends counterfactually on its cause within the contextually relevant possibility horizon. As demonstrated by cases of redundant causation, however,

---

[20] See e.g. Lewis (1986*g*), p. 259; Sartorio (2010), p. 261.

counterfactual dependence is not a necessary condition for causation. To find a necessary condition that captures the sense in which a cause must make a difference to its effect, we therefore need a weaker version of counterfactual dependence: as I propose in the following, we need the doubly modal relation of security-dependence.

3. Security-dependence

My aim in this section is to define the relation of security-dependence. I begin by defining the modal notion of *security within a possibility horizon* (section 3.1). Building on this, I then define the doubly modal relation of *security-dependence within a possibility horizon* (section 3.2).

3.1 Security

There is a range of expressions that convey the idea that an event could very easily have failed to occur. We may say, for example, that we came through something 'by the skin of our teeth', that an outcome 'was hanging by a thread', that we came 'within a hair's breadth' of another outcome, etc. Such cases may be contrasted with cases where an event could *not* easily have failed to occur – though we do not have quite so colourful expressions to describe these cases. My aim in introducing the notion of *security* is to capture this intuitive sense in which an event could more or less easily have failed to occur.

Let e be an instantaneous event that occurs at time t' in @, and let t be a time that is strictly earlier than time t'. Intuitively, we may say that the circumstances at time t were such that e could easily have failed to occur, when even slight changes to the circumstances at time t would be sufficient to prevent e's occurrence. And we may say that the circumstances at time t were such that e could *not* easily have failed to occur, when only major changes to the circumstances at time t could have prevented e's occurrence.

Our task now is to make this precise in a way that does not draw on implicitly or explicitly causal notions. To do so, I will rely on the notion of a possibility horizon. Thus, my aim is to define the notion of *security within a possibility horizon*.

A first suggestion is that *e*'s security at an earlier time *t* and within a possibility horizon $\mathscr{H}$ is given by the distance-at-*t* from @ to the closest-at-*t* world(s) in $\mathscr{H}$ where *e* does not occur. When there is only a short distance-at-*t* from @ to the closest-at-*t* world(s) in $\mathscr{H}$ where *e* does not occur, *e* has a very low degree of security at *t* within $\mathscr{H}$: just a slight change to the circumstances at time *t* takes us from @ to a world within $\mathscr{H}$ where *e* does not occur. By contrast, when the closest-at-*t* world(s) in $\mathscr{H}$ where *e* does not occur are very distant-at-*t*, *e* has a high degree of security at *t* within $\mathscr{H}$: it would require major changes to the circumstances at time *t* to take us from @ to a world within $\mathscr{H}$ where *e* does not occur.

This first suggestion goes a long way towards capturing our intuitive notion of how easily an event could have failed to occur, and in many of the cases we shall consider in the following, it is fully adequate. In some cases, however, there are several different possible changes to the circumstances at time *t*, each of which would be sufficient to prevent the later occurrence of *e*. To illustrate, consider our standard case of causation by omission:

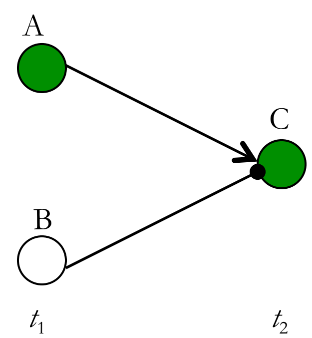


Figure 1

Let us treat the world of the example as actual, and suppose that the contextually relevant possibility horizon $\mathscr{H}$ includes the following four possible worlds (cf. section 1.1):

	Complete state at time t_1:	*Events at t_1:*		t_2:
@:	**A** fires, **B** does not fire	A,	$\neg B$,	C
w_1:	**A** does not fire, **B** does not fire	A,	$\neg B$,	C
w_2:	**A** fires, **B** fires	A,	$\neg B$,	C
w_3:	**A** does not fire, **B** fires	A,	$\neg B$,	C

In this case, there are two different possible changes to the circumstances at time t_1, both of which would prevent the occurrence of C: if A had not occurred, C would not have occurred; and if $\neg B$ had not occurred, C would not have occurred.

Intuitively, both of these threats should play a role in our evaluation of C's security at time t_1. However, the simple suggestion presented above can only take account of one of them – namely, the threat that is realized in the closest-at-t_1 world where C does not occur. To see this, suppose for example that the closest-at-t_1 world in $\mathscr{H}$ where C does not occur is w_1. In that case, the simple suggestion presented above yields the result that C's security at time t_1 and within possibility horizon $\mathscr{H}$ is given by the distance-at-t_1 from @ to w_1. However, this does not at all take into account that there is also a second change to the circumstances at time t_1, which would prevent C's occurrence – namely, the change consisting in letting neuron **B** fire.

To take account of multiple threats, as in the case illustrated here, we need a slightly more complex notion of security. In brief, my suggestion is that we should evaluate the security of an event e at an earlier time t and within a possibility horizon $\mathscr{H}$ as follows: first, we determine all the ways in which circumstances could have been different at time t within possibility horizon $\mathscr{H}$, such that, had they been different in that way, e would not have occurred. Second, we assess the distance-at-t to each of the closest-at-t world(s) where things are different in one of these ways (and where e therefore does not occur). Thus, the security of an event is a measure of the totality of these distances.

To make this suggestion precise, I will rely on the notion of a *minimal dependence set*.[21] Let us first define the more basic notion of a dependence set:

> *Dependence set*: a set S of contemporaneous events is a *dependence set* for e within possibility horizon $\mathscr{H}$ iff:
>
> a) S occurs at a time t strictly earlier than e,
>
> b) there is at least one world in $\mathscr{H}$ where *none* of the events in S occur, and
>
> c) in the closest-at-t worlds in $\mathscr{H}$ where *none* of the events in S occur, e does not occur.

Note that this notion of a dependence set simply generalises my above definition of counterfactual dependence to *sets* of contemporaneous events.

In the example above, we find that $\{A\}$, $\{\neg B\}$, and $\{A, \neg B\}$ are all dependence sets for C at t_1 within possibility horizon $\mathscr{H}$. There is a clear sense in which these dependence sets point to possible threats to C: if A had failed to occur, C would not have occurred; if $\neg B$ had failed to occur, C would not have occurred; and if A and $\neg B$ had both failed to occur, C would not have occurred. However, these threats are not properly independent of each other: once we have taken account of the threat that A fails to occur, and the threat that $\neg B$ fails to occur, we do not need to also take account of the threat that A and $\neg B$ both fail to occur.

To address this problem, we need the notion of a *minimal* dependence set. To define this, the following notation will be useful:

> For any set S of events, $w(—S)$ is the set of worlds such that *none* of the events in S occur.[22]

[21] A similar notion of a minimal dependence set was introduced by Ramachandran (1997), p. 270, who aimed to use it to handle cases of preemption.

[22] Note that this definition requires us to quantify over all metaphysically possible worlds, with absolutely no restrictions on our domain of quantification. It does not merely require us to quantify over nomologically possible worlds, or worlds that are included in the contextually relevant possibility horizon.

Note that this definition concerns *all* possible worlds, not just nomologically possible worlds or worlds that are included in the relevant possibility horizon. On this basis, we may now define a *minimal* dependence set as follows, drawing on a notion of minimality that is clearly parallel to the notion of minimality that I used in my definition of minimal sufficiency.

> *Minimal dependence set*: a set S of contemporaneous events is a *minimal dependence set* for e within possibility horizon $\mathscr{H}$ iff:
>
> a) S is a dependence set for e within $\mathscr{H}$, and
>
> b) there is no set of contemporaneous events S^*, such that
>
> > i) S^* is a dependence set for e within $\mathscr{H}$,
> >
> > ii) S and S^* have the same realization in @, and
> >
> > iii) $w(—S) \subset w(—S^*)$.

Note that in the special case of an event e that occurs in *every* world within a possibility $\mathscr{H}$, there are no dependence sets, and hence no minimal dependence sets either. This corresponds nicely to the intuition that in this special case there are *no possible threats* to e within the possibility horizon $\mathscr{H}$.

In our case of causation by omission, we now find that C has two minimal dependence sets at t_1 within $\mathscr{H}$ – namely, $\{A\}$ and $\{\neg B\}$. By contrast, $\{A, \neg B\}$ is not a *minimal* dependence set for C within $\mathscr{H}$, since the set of worlds where both A and $\neg B$ fail to occur is a proper subset of the set of worlds where A does not occur. These minimal dependence sets capture the threats to C without unwanted redundancy.

The closest-at-t_1 world in $\mathscr{H}$ where A fails to occur is w_1, and the closest-at-t_1 world in $\mathscr{H}$ where $\neg B$ fails to occur is w_2. My suggestion therefore is that C's security at t_1 within $\mathscr{H}$ is given by the distance-at-t_1 from @ to w_1 together with the distance-at-t_1 from @ to w_2.

More generally, we may define security within a possibility horizon as follows:

> *Security*: let e be an instantaneous event that occurs at time t', and let t be a time that is strictly earlier than t'. Let $S_1, S_2, \ldots, S_n$ be sets of contemporaneous events occurring at t, such that $S_1, S_2, \ldots, S_n$ are the minimal dependence sets for e within possibility horizon $\mathscr{H}$.
>
> Then the security of e at t within $\mathscr{H}$ is given by the distance-at-t, for each set S_i, between @ and the closest-at-t world(s) in $\mathscr{H}$ where *none* of the events in S_i occur.

Note that so far we have only been concerned with the security of *actual* events. However, the definition of security may be straightforwardly extended to merely possible events, i.e. events that do not occur in the actual world @. Suppose that e is a merely possible event. On our first simple understanding of security, e's security at a time t in @ is given by the distance-at-t from @ to the closest-at-t world where e does not occur. But the closest-at-t world where e does not occur is @ itself, and so the relevant distance is zero.

The definition in terms of minimal dependence sets may be extended to give the same result: we may say that a merely possible event e has just one minimal dependence set – namely, the empty set Ø. From the standpoint of @, the closest-at-t world where *none* of the events in Ø occur is @ itself. And, once again, the distance from @ to itself is zero. Thus we may extend the definition of security as follows:

> *Minimal security*: an event e has *minimal security* at t within a possibility horizon $\mathscr{H}$ iff e does not occur in @.

At the other extreme, we shall say that an event e has *maximal* security within a possibility horizon $\mathscr{H}$ just in case e occurs in *every* world within $\mathscr{H}$:

> *Maximal security*: an event e has *maximal security* within a possibility horizon $\mathscr{H}$ iff e occurs in every world in $\mathscr{H}$.

In the intermediate cases, where an event e occurs, but does not occur in every world within the relevant possibility horizon $\mathscr{H}$, the security of e within $\mathscr{H}$ is neither minimal nor maximal, but somewhere in between.

In the following, all we need are facts about *comparative* security. In making judgements about comparative security, we may rely on the following two *ceteris paribus* rules when making judgements of *comparative* security:

The first is that, *ceteris paribus*, an event is more secure when there is a greater distance-at-t to the closest-at-t world(s) where none of the events in its minimal dependence set(s) occur. To take a simple example, suppose that the security of e at t within $\mathscr{H}$ is given by the distance-at-t between @ and w_1, and suppose that the security of e^* at t within $\mathscr{H}$ is given by the distance-at-t between @ and w_2. In that case, we can settle the question of comparative security simply by comparing distances: e is more secure at t than e^* just in case the distance-at-t from @ to w_1 is strictly greater than the distance-at-t from @ to w_2, and e is just as secure at t as e^* just in case the distance-at-t from @ to w_1 is exactly the same as the distance-at-t from @ to w_2.

The second rule is that, *ceteris paribus*, an event is more secure when it has fewer minimal dependence sets. Suppose, for example, that the security of e at t within $\mathscr{H}$ is given by the distance-at-t from @ to w_1, and that the security of e^* at t within $\mathscr{H}$ is given by the distance-at-t from @ to w_1 together with the distance-at-t from @ to w_2. In that case, e is more secure at t than e^*.

This notion of security within a possibility horizon gives a deterministic interpretation of our intuitive judgements about how easily an event could have failed to occur, based on the circumstances at some earlier time: saying that an event e has a high degree of security at a time t captures the idea that e could not easily have failed to occur, based on the circumstances at t. And saying that an event e has a low degree of security at an earlier time t captures the idea that *e could* easily have failed to occur, based on the circumstances at t. In the following, I will now build on this to define the relation of security-dependence.

3.2 Security-dependence

Given the notion of an event's security within a possibility horizon, we may now go on to consider how an event *c influences* the security of a later event *e*, where we shall understand influence in counterfactual terms. In particular, considering the time *t* at which *c* occurs, we may ask: how would *e*'s security at *t* have been different, if *c* had not occurred?

We may distinguish several possible answers to this question. First, we need to distinguish between cases where the relevant possibility horizon $\mathscr{H}$ is such that *c* occurs in every world within $\mathscr{H}$, and cases where $\mathscr{H}$ contains at least one world where *c* does not occur.

If *c* occurs in every world in $\mathscr{H}$, then any counterfactual claim with the antecedent 'if *c* had not occurred, then . . .' is false within $\mathscr{H}$, since $\mathscr{H}$ contains no world in which the antecedent is satisfied.

If $\mathscr{H}$ does contain at least one not-*c*-world, we may, by contrast, get true counterfactuals with the antecedent 'if *c* had not occurred, then . . .'. In particular, we may get true counterfactuals comparing *e*'s actual security at time *t* with the security at *t* that *e* would have had, if *c* had not occurred. Such counterfactuals require us to evaluate the security of *e* in the closest-at-*t* not-*c*-worlds in $\mathscr{H}$. For any such world *w*, this is done exactly as described in section 3.1, by taking the standpoint of world *w* (cf. footnote 8). And we may then compare *e*'s security at *t* in the actual world @ with the security at *t* that *e* would have had, if *c* had not occurred. In this comparison, we may distinguish several possibilities.

In some cases, the closest-at-*t* not-*c*-world(s) within a possibility horizon $\mathscr{H}$ do not agree on whether *e* is *less secure* at *t*, *just as secure* at *t*, or *more secure* at *t*. In such cases, we typically resort to contrastive causal claims, which artificially restrict the possibility horizon under consideration.[23] I discuss such cases in more detail in Chapter 10 section 3.

---

[23] Cf. Northcott (2008), p. 120.

In most cases, however, all the closest-at-*t* not-*c*-worlds in $\mathscr{H}$ do agree on whether *e* is *less secure* at *t*, *just as secure* at *t*, or *more secure* at *t*. In such uniform cases, we may distinguish the following three relations:

Security-dependence: an actual event *e security-depends* on an actual event *c* within possibility horizon $\mathscr{H}$ iff

a) *c* occurs at a time *t* strictly earlier than *e*,
b) there is at least one not-*c*-world in $\mathscr{H}$, and
c) in the closest-at-*t* not-*c*-world(s) in $\mathscr{H}$, *e* is *less secure* at *t*.

Security-independence: an actual event *e* is *security-independent* from an actual event *c* within possibility horizon $\mathscr{H}$ iff

a) *c* occurs at a time *t* strictly earlier than *e*,
b) there is at least one not-*c*-world in $\mathscr{H}$, and
c) in the closest-at-*t* not-*c*-world(s) in $\mathscr{H}$, *e* is *just as secure* at *t*.

Negative security-dependence: an actual event *e security-depends negatively* on an actual event *c* within possibility horizon $\mathscr{H}$ iff

a) *c* occurs at a time *t* strictly earlier than *e*,
b) there is at least one not-*c*-world in $\mathscr{H}$, and
c) in the closest-at-*t* not-*c*-world(s) in $\mathscr{H}$, *e* is *more secure* at *t*.

It is important to note that these three relations are doubly modal: the notion of security is itself a modal notion – to evaluate an event *e*'s security at a time *t*, we need to consider the minimal dependence sets for *e* within the relevant possibility horizon $\mathscr{H}$ and the distance-at-*t*, for each of these sets, to the closest-at-*t* world(s) in $\mathscr{H}$ where all events in this set fail to occur. To evaluate which of the three relations holds between an event *c*, occurring at *t*, and a later event *e*, we then need to compare *e*'s *actual* security at *t* with the security *e would have had* at *t*, if *c* had not occurred.

My suggestion is that the distinction between these three different relations of security-dependence, security-independence, and negative security-dependence is metaphysically significant. In particular, I suggest that security-dependence within a possibility horizon is a necessary condition for causation:

> *The second condition – security-dependence*:
> c is a cause of e within a possibility horizon $\mathscr{H}$
> *only if* e security-depends on c within $\mathscr{H}$.

In Appendix A section 4, I prove the important result that counterfactual dependence within a possibility horizon is sufficient for security-dependence within that same possibility horizon. Thus, we find that the condition of security-dependence correctly captures all paradigm cases of difference-making: in all paradigm cases of difference-making, where the effect depends counterfactually on its cause(s) within the contextually relevant possibility horizon, we also find that the effect security-depends on its cause(s) within the contextually relevant possibility horizon.

In Chapter 9 section 1, we shall see that the effect also security-depends on its cause(s) in cases of redundant causation, showing that security-dependence is a strictly weaker version of counterfactual dependence. In this way, security-dependence captures the sense in which a cause *must* make a difference to its effect.

Furthermore, I suggest that the distinction between the three relations of security-dependence, security-independence, and negative security-dependence allows us to capture an important distinction between different kinds of causal relevance. The notion of causal relevance is most easily understood within an indeterministic framework. Within such a framework, we may say that the occurrence of c has *positive causal relevance* for the occurrence of e just in case c's occurrence raises the probability of e's occurrence; the occurrence of c is *causally irrelevant* for the occurrence of e just in case c's occurrence leaves the probability of *e*'s occurrence unchanged; and the occurrence of *c* has *negative causal relevance* for the occurrence of *e* just in case *c*'s occurrence lowers the probability of *e*'s occurrence. As further elucidation, we tend to say that *c* is *irrelevant* to *e* when *c*'s occurrence is causally irrelevant to *e*'s occurrence, and we tend to say that *e* occurs *in spite of c* when *c*'s occurrence has negative causal relevance for the occurrence of *e*.

Relying on the distinction between security-dependence, security-independence, and negative security-dependence, we may now capture the distinction between these three different kinds of causal relevance within a deterministic framework:

> *Positive causal relevance*: *c* has positive causal relevance to *e* within $\mathscr{H}$ iff *e* security-depends positively on *c* within $\mathscr{H}$.
>
> *Causal irrelevance*: *c* is causally irrelevant to *e* within $\mathscr{H}$ iff *e* is security-independent of *c* within $\mathscr{H}$.
>
> *Negative causal relevance*: *c* has negative causal relevance to *e* within $\mathscr{H}$ iff *e* security-depends negatively on *c* within $\mathscr{H}$.

4. Conclusion

In this chapter, I have presented a second necessary condition for causation – the condition of security-dependence. The motivation for this condition is the intuition that a cause must somehow *make a difference* to its effect. The condition of security-dependence, which is strictly weaker than counterfactual dependence, formulates a necessary condition for causation that captures this intuition.

In the following two chapters, we shall see how accepting security-dependence as a necessary condition for causation yields intuitively correct results in a wide variety of cases. We shall furthermore see how recognising the two relations of security-independence and negative security-dependence allows us to capture the nuances of our intuitive judgements in cases where we judge that an event *c* is *not* a cause of a later event *e*: in these cases, the relations of security-independence and negative security-dependence correctly capture whether we judge that *c* is simply *irrelevant* to *e*, or that *e* occurs *in spite of c*.

9

Applications of security-dependence

In this chapter, we shall see some of the central applications of my proposal that security-dependence is a necessary condition for causation. I first show that the effect security-depends on its cause(s) in cases of redundant causation (section 1). Next, I show how security-dependence allows my account of causation to accommodate our intuitive judgements on the counterexamples to the transitivity and intrinsicness of causation (section 2 and 3).

1. Redundant causation

In this section, I apply the condition of security-dependence to cases of redundant causation. The characteristic feature of these cases is that the effect does not depend counterfactually on its cause(s). For this reason, cases of redundant causation present a challenge when we are trying to spell out the sense in which a cause *must* make a difference to its effect. As we shall see in the following, the condition of security-dependence allows us to overcome this challenge: it captures a sense in which the cause makes a difference to its effect, even in cases where the effect does not depend counterfactually on its cause(s).

To show this, I first show that the effect security-depends on its cause in our standard case of early preemption, and note that precisely the same reasoning applies to our standard cases of late preemption and symmetric overdetermination (section 1.1). I next discuss our slightly more complex case of trumping preemption, showing that the effect security-depends on its cause within each of the two possibility horizons that can reasonably be used to judge this case (section 1.2).

1.1 Early preemption, late preemption, and symmetric overdetermination

Let us begin by considering our standard case of early preemption, illustrated in the neuron diagram below:

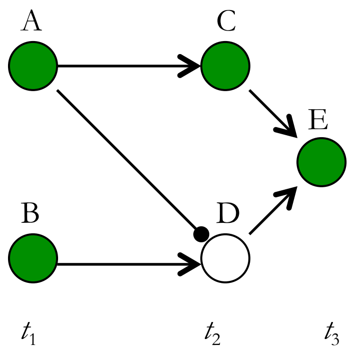


Figure 1

Intuitively, *A* is a cause of *E*. Does *E* security-depend on *A* within the contextually relevant possibility horizon?

It does. To show this, our first task is to identify the relevant possibility horizon. Considering the time at which *A* occurs, and following the standard practice of treating neuron firings as deviant events, we arrive at a possibility horizon containing the following four worlds, characterised by their complete state at time t_1:

	Complete state at time t_1:	*Events at t_1:*	t_3:
@:	**A** fires, **B** fires	*A*, *B*	*E*
w_1:	**A** does not fire, **B** fires	*A*, *B*	*E*
w_2:	**A** fires, **B** does not fire	*A*, *B*	*E*
w_3:	**A** does not fire, **B** does not fire	*A*, *B*	*E*

Call this possibility horizon $\mathscr{H}$. The relevant relations of comparative distance-at-t_1 are represented in Figure 2 below. Note that the worlds where *E* occurs – namely, @, w_1, and w_2 – are white, while the world where *E* does not occur – namely, w_3 – is coloured grey:

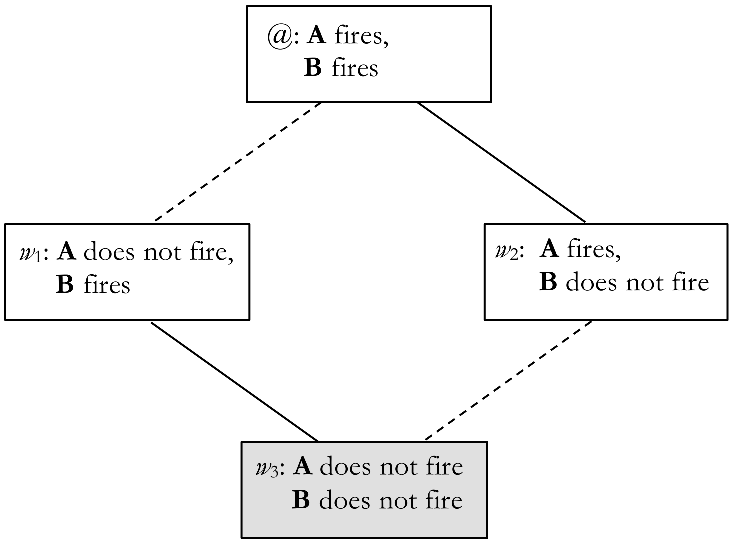


Figure 2

To evaluate whether E security-depends on A within $\mathscr{H}$, we now proceed in the following three steps. Note that I leave out relativisations to the possibility horizon $\mathscr{H}$ in order to avoid clutter:

Step 1: the first step is to evaluate E's security at t_1 in @. From the standpoint of @, there is just one minimal dependence set for E at time t_1 – namely, $\{A, B\}$. The closest-at-t_1 world where both A and B fail to occur is w_3. Thus, E's security at t_1 in @ is given by the distance-at-t_1 from @ to w_3.

Step 2: the second step is to evaluate E's security at t_1 in the closest-at-t_1 world(s) where A does not occur. From the standpoint of @, the closest-at-t_1 world where A does not occur is w_1. From the standpoint of w_1, there is one minimal dependence set for E at time t_1 – namely, $\{B\}$; and the closest-at-t_1 world where B fails to occur is w_3. Thus, E's security at t_1 in w_1 is given by the distance-at-t_1 from w_1 to w_3.

Step 3: the final step is to compare E's security at t_1 in @ with E's security at t_1 in w_1. Since the distance-at-t_1 from w_1 to w_3 is strictly smaller than the distance-at-t_1 from @ to w_3, we find that E is *less secure* at t_1 in w_1 than it is at t_1 in @.

Thus we find that *E* security-depends on *A* within our possibility horizon $\mathscr{H}$. Even though *E* does not depend counterfactually on *A*, the condition of security-dependence is thus satisfied.[1] We can – by applying exactly the same three-step procedure – verify that the effect also security-depends on its cause(s) in standard cases of late preemption and symmetric overdetermination.

Our case of trumping preemption is slightly more complex. For the sake of illustration, I therefore go through this case in the following section.

1.2 A variant of trumping preemption

In this section, I show that the effect also security-depends on its cause in our case of trumping preemption. For ease of reference, I repeat the neuron diagram here (for a full description of the case, see Chapter 7 section 2.3):[2]

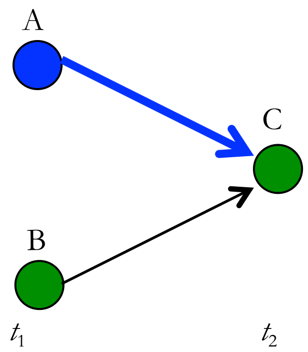


Figure 3

Intuitively, *A-blue* causes *C*. Correspondingly, we find that *C* security-depends on *A-blue* within any reasonable choice of possibility horizon. To see this, note that there are two reasonable choices of possibility horizon in the present case.

[1] Note that an application of the three steps also yields the result that *E* security-depends on *B*. Thus, the condition of security-dependence cannot distinguish between genuine causes and preempted backups. This result presents no threat to my claim that security-dependence is a necessary condition for causation. It simply shows that security-dependence is not *sufficient* for causation – we need the additional condition of process-connection to distinguish genuine causes from preempted backups.

[2] Cf. Paul and Hall (2013), Figure 26.

Let us begin by considering the larger possibility horizon $\mathscr{H}_1$, which includes all of the following worlds:

	Complete state at time t_1:	*Events at t_1:*		*t_2:*
@:	**A** fires in blue, with intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_1:	**A** fires not in blue, with intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_2:	**A** fires in blue, without intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_3:	**A** fires not in blue, without intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_4:	**A** does not fire; **B** fires	*A-blue*,	*B*	*C*
w_5:	**A** fires in blue, with intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_6:	**A** fires not in blue, with intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_7:	**A** fires in blue, without intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_8:	**A** fires not in blue, without intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_9:	**A** does not fire; **B** does not fire	*A-blue*,	*B*	*C*

Within this possibility horizon, we simply find that *C* depends counterfactually on *A-blue*: from the standpoint of @, the closest-at-t_1 world where *A-blue* does not occur is world w_1, where **A** still fires with intensity *I*, but not in blue. *C* does not occur in w_1. Since counterfactual dependence within a possibility horizon entails security-dependence within that same possibility horizon, we thus straightforwardly find that *C* security-depends on *A-blue* within possibility horizon $\mathscr{H}_1$.

The second reasonable choice of possibility horizon derives from a context where the only alternative to **A**'s firing in blue and with intensity *I* is that **A** does not fire at all. This yields the following more restricted possibility horizon $\mathscr{H}_2$:

	Complete state at time t_1:	*Events at t_1:*		*t_2:*
@:	**A** fires in blue, with intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_4:	**A** does not fire; **B** fires	*A-blue*,	*B*	*C*
w_5:	**A** fires in blue, with intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_9:	**A** does not fire; **B** does not fire	*A-blue*,	*B*	*C*

It is within this more restricted possibility horizon that the case earns its title as a case of redundant causation: within this more restricted possibility horizon, we find that *C* does not depend counterfactually on *A-blue*: from the standpoint of @, the closest-at-t_1 world where *A-blue* does not occur is w_4. And *C* still occurs in w_4 – because of *B*.

However, we still find that *C* security-depends on *A-blue* within $\mathscr{H}_2$. Indeed, the argument to show this is entirely parallel to the argument showing that the effect security-depends on its cause in our standard case of early preemption considered above. To bring out this parallel, I illustrate the comparative distances-at-t_1 in the figure below, where the world where *C* does not occur – namely, w_9 – is coloured grey:

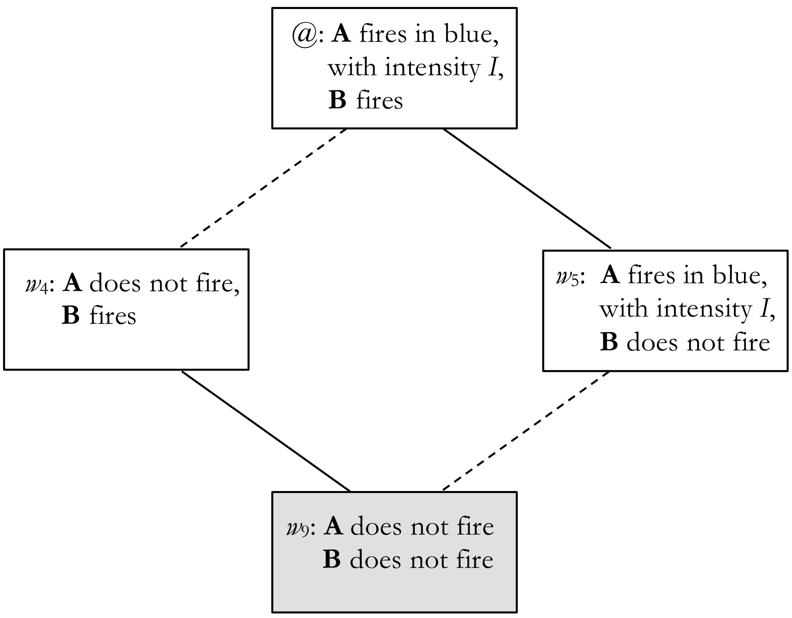


Figure 4

By exactly the same argument as above, we now find – as we should – that *C* security-depends on *A-blue* within $\mathscr{H}_2$.

In this way, the condition of security-dependence captures how a cause makes a difference to its effect even in cases of redundant causation, where the effect does not depend counterfactually on its cause(s).

2. The non-transitivity of causation

In this section, I show how accepting security-dependence as a necessary condition for causation allows my account to accommodate our intuitive judgements on counterexamples to the transitivity of causation. I begin by outlining how transitivity fails in these cases (section 2.1). I then turn to the counterexamples themselves.

Counterexamples to the transitivity of causation are standardly divided into two groups: counterexamples that do not involve omissions, and counterexamples that do.[3] We have already seen examples of both kinds: *Switch* (discussed in Chapter 2 section 2.2.1) does not involve omissions, while *Boulder* (discussed in Chapter 2 section 1.3.2) does involve omissions. In the following, I show how the condition of security-dependence secures the intuitively correct verdict on both *Switch* (section 2.2) and *Boulder* (section 2.3). In section 2.4, I argue that – in keeping with my general claim that there is no deep metaphysical distinction between omissions and ordinary events – counterexamples to the transitivity of causation should not be categorised by whether or not they involve omissions; rather, they should be categorised by *how* the condition of security-dependence fails to be satisfied.

In the two final sections, I discuss how security-dependence may be understood as drawing a line between preemption cases and transitivity failure cases (section 2.5), and I consider an apparent counterexample to my claim that security-dependence allows us to handle *all* transitivity failure cases (section 2.6).

[3] See e.g. Paul and Hall (2013), pp. 215-237.

2.1 How transitivity fails

The intuitive principle that causation is a transitive relation may be stated as follows (the asterisk signifies that I do not endorse this principle):

> **Transitivity of causation*: if there is a set of events $\{d_1, d_2, \ldots, d_n\}$, such that c causes d_1, d_1 causes d_2, . . . , and d_n causes e, then c causes e.

We have already seen that *Transitivity of causation* faces counterexamples: there are cases where causation *fails to behave transitively*, i.e. cases where c does not cause e, even though there is a set of events $\{d_1, d_2, \ldots, d_n\}$, such that c causes d_1, d_1 causes d_2, . . . , and d_n causes e.

The condition of process-connection cannot on its own accommodate our intuitive verdicts in these cases: as I prove in Appendix A section 2, the relation of process-connection is transitive. To accommodate our intuitive verdicts and explain why causation fails to behave transitively in these cases, we therefore need to supplement the condition of process-connection with a further necessary condition, which is *not* transitive.

This is where the condition of security-dependence comes in: security-dependence is *not* a transitive relation. Let us say that *security-dependence fails to behave transitively* whenever e does not security-depend on c, even though there is a set of events $\{d_1, d_2, \ldots, d_n\}$, such that d_1 security-depends on c, d_2 security-depends on d_1, . . . , and e security-depends on d_n. My suggestion is that causation fails to behave transitively in exactly those cases where security-dependence fails to behave transitively. And the fact that security-dependence fails to behave transitively in these cases *explains* why causation fails to behave transitively: in the relevant cases, there is a set of events $\{d_1, d_2, \ldots, d_n\}$, such that c causes d_1 (and correspondingly, d_1 security-depends on c), d_1 causes d_2 (and correspondingly, d_2 security-depends on d_1), . . . , and d_n causes e (and correspondingly, e security-depends on d_n), but e does not security-depend on c – and for this reason, c is not a cause of e.

2.2 Switch

Let us begin by considering a counterexample to *Transitivity of causation* that does not involve omissions, namely *Switch*:

> *Switch*: Suzy is standing by a switch in the railroad tracks. She sees a train approaching in the distance, and flips the switch so that the train travels down the left-hand track. If she had not flipped the switch, the train would have travelled down the right-hand track instead. Since the tracks converge a few miles later, the train arrives at its destination all the same.[4]

The structure of *Switch* is represented in the neuron diagram below (for details, see Chapter 2 section 2.2.1):[5]

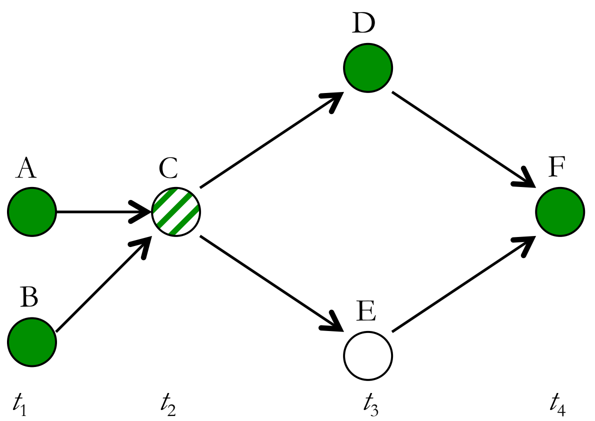


Figure 5

Switch presents a counterexample to *Transitivity of causation* as follows: *B* (Suzy's flipping the switch) causes *D* (the train's journey down the left-hand track), and *D* in turn causes *F* (the train's arrival at its destination). By transitivity, it follows that *B* causes *F*. But, intuitively, this is false: intuitively, *B* (Suzy's

[4] Switching cases of this kind are discussed e.g. in Hall (2004*a*), pp. 187-92, (2007*a*), p. 118, and (2007*b*), pp. 28-32 and 51; Hitchcock (2009), p. 394; Mackie (1992); p. 496-7; Paul and Hall (2013), p. 232; and Sartorio (2005), pp. 74-5.
[5] The figure is adapted from Paul and Hall (2013), p. 232.

flipping the switch) does *not* cause *F* (the train's arrival at its destination) – rather, *B* is simply irrelevant to *F*.[6]

The condition of process-connection cannot on its own accommodate this verdict: as is easily verified, *B* is process-connected to *D*, and *D* is in turn process-connected to *F*. By the transitivity of process-connection (cf. Appendix A section 2), it follows that *B* is process-connected to *F* – as indeed it is, via the following process:

t_1:	*B*	(together with *A*)
t_2:	*C-stripes*	
t_3:	*D*	
t_4:	*F*	

This process is illustrated in the figure below, where I have put a halo around the neurons that are part of the process:

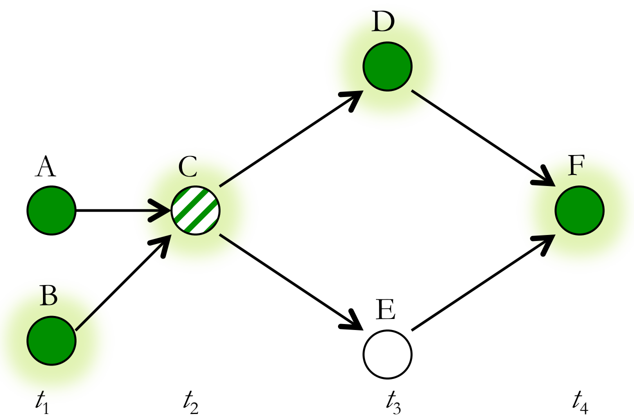


Figure 5*

Thus, the condition of process-connection cannot on its own accommodate our intuitive verdict that *B* does not cause *F*. To accommodate this verdict – and to explain why causation fails to behave transitively in *Switch* – we need the

[6] For this judgement, see e.g. Hall (2007*b*), pp. 28-29; Hitchcock (2009), p. 394; and Paul and Hall (2013), p. 232.

condition of security-dependence: as we shall now see, the reason why causation fails to behave transitively in *Switch* is that security-dependence fails to behave transitively. In brief, *F* security-depends on *D*, and *D* security-depends on *B*, but *F* does not security-depend on *B* – and this is the reason why *B* is not a cause of *F*.

It is easy to see that *F* security-depends on *D*: within any possibility horizon that treats *D* as a candidate cause, *F* depends counterfactually on *D*.[7] Since counterfactual dependence within a possibility horizon implies security-dependence within that same possibility horizon, this shows that *F* security-depends on *D*. It is similarly easy to see that *D* security-depends on *B*: once again, *D* depends counterfactually on *B* within any possibility horizon that treats *B* as a candidate cause.

I will now show that *F* does *not* security-depend on *B* within the contextually relevant possibility horizon. Note that our possibility horizon should include the following four possible worlds, characterised by their complete state at time t_1:

	Complete state at time t_1:	*Events at* t_1:	t_4:
@:	**A** fires, **B** fires	*A*, *B*	*F*
w_1:	**A** does not fire, **B** fires	*A*, *B*	*F*
w_2:	**A** fires, **B** does not fire	*A*, *B*	*F*
w_3:	**A** does not fire, **B** does not fire	*A*, *B*	*F*

Call this possibility horizon $\mathscr{H}$. We may represent the comparative distances-at-t_1 between these four worlds as follows, where the worlds in which *F* does not occur are coloured grey:

[7] Remember that my proposed recipe for evaluating counterfactuals ensures that the relevant counterfactuals receive a non-backtracking reading.

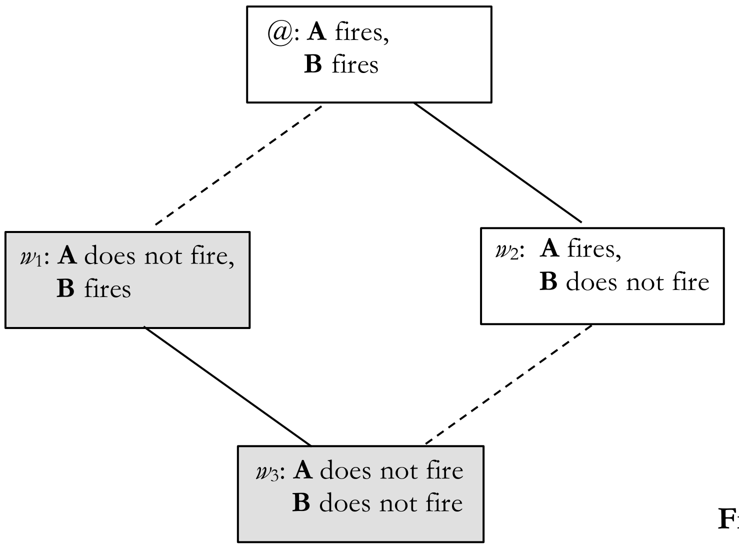


Figure 6

To see that F does not security-depend on B within $\mathscr{H}$, we proceed as follows:

Step 1: our first step is to evaluate F's security at t_1 in @. From the standpoint of @, there is just one minimal dependence set for F at t_1 – namely, $\{A\}$; and the closest-at-t_1 world where A fails to occur is w_1. Thus, F's security at t_1 in @ is given by the distance-at-t_1 from @ to w_1.

Step 2: our second step is to evaluate F's security at t_1 in the closest-at-t_1 world(s) where B does not occur. From the standpoint of @, the closest-at-t_1 world where B does not occur is w_2. From the standpoint of w_2, there is just one minimal dependence set for F at t_1 – namely, $\{A\}$; and the closest-at-t_1 world where A fails to occur is world w_3. Thus, F's security at t_1 in w_2 is given by the distance-at-t_1 from w_2 to w_3.

Step 3: our final step is to compare F's security at t_1 in @ with F's security at t_1 in w_2. As indicated in the figure, the distance-at-t_1 from @ to w_1 is exactly the same as the distance-at-t_1 from w_2 to w_3. This shows that F is *just as secure* at t_1 in w_2 as it is in @.

Thus, F does not security-depend on B within $\mathscr{H}$.[8] And for this reason, B is not a cause of F: the condition of security-dependence fails to be satisfied.

---

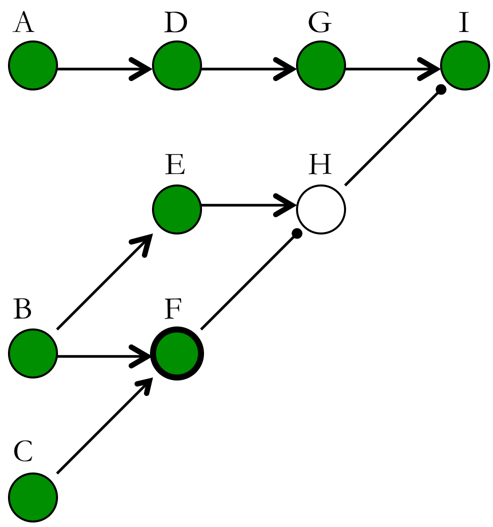
[8] Within the smaller possibility horizon where A is treated as a background condition, while B is treated as a candidate cause, we similarly find that F is security-independent from B.

2.3 Boulder

The condition of security-dependence similarly yields intuitively correct results when applied to the following omission-involving case (cf. Chapter 2 section 1.3.2):

> *Boulder*: 'A boulder is dislodged and begins rolling ominously toward Hiker. Before it reaches him, Hiker sees the boulder and ducks. The boulder sails harmlessly over his head with nary a centimetre to spare. Hiker survives his ordeal.'[9]

The structure of this case is represented below:[10]



t_1 t_2 t_3 t_4

Figure 7

A here represents the physiological processes that keep Hiker alive, *B* represents the boulder's fall, and *C* represents Hiker's being attentive. *E* represents the boulder's flying through the air, *F* represents Hiker's duck, ¬*H* represents Hiker's not being hit by the boulder, and *I* represents Hiker's survival. Note that **F** is a stubborn neuron (marked by its thicker perimeter),

[9] Paul and Hall (2013), p. 222. The case is originally due to Hall.

[10] This is a modified version of Paul and Hall (2013), Figure 10.

requiring two stimulatory signals to fire. This reflects the fact that Hiker will duck just in case the boulder falls *and* he is attentive.

Boulder presents a counterexample to *Transitivity of causation* as follows: *B* (the boulder's fall) causes *F* (Hiker's duck), and *F* in turn causes *I* (Hiker's survival). By transitivity, it follows that *B* causes *I*. But this is false: *B* (the boulder's fall) does *not* cause *I* (Hiker's survival) – rather, Hiker survives *in spite of* the boulder's fall.

As in *Switch*, process-connection cannot on its own accommodate our intuitive judgement in this case: *B* is process-connected to *F*, and *F* is process-connected to *I*. By the transitivity of process-connection, it follows that *B* is process-connected to *I* – and indeed it is, via the process:

t_1:	B	(together with C)
t_2:	F	
t_3:	$\neg H$	(together with G)
t_4:	I	

This process is illustrated below:

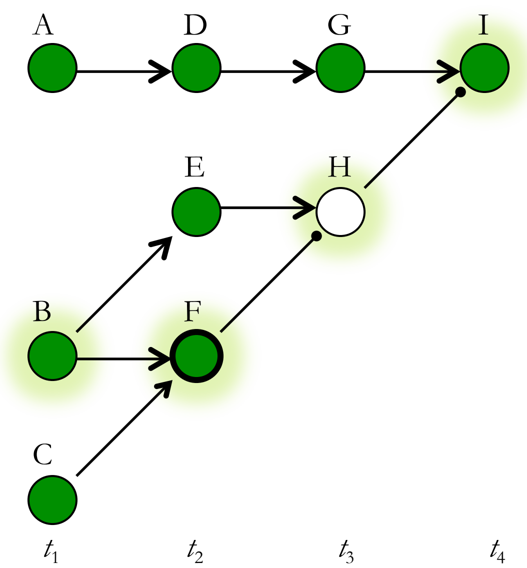


Figure 7*

This shows that the condition of process-connection cannot explain why causation fails to behave transitively in this case. To explain this – and to accommodate the intuitive verdict that B does not cause I – we need the further condition of security-dependence: as we shall now see, the fact that causation fails to behave transitively in *Boulder* is explained, as in *Switch*, by the fact that security-dependence fails to behave transitively – I security-depends on F, and F security-depends on B, but I does not security-depend on B.

This is seen as follows: I security-depends on F, since I depends counterfactually on F within any possibility horizon that treats F as a candidate cause. And F security-depends on B, since there is once again counterfactual dependence within any possibility horizon that treats B as a candidate cause. However, I does not security-depend on B within the contextually relevant possibility horizon.

To see this, note that it is natural to treat A (the physiological processes that keep Hiker alive) as a default event. Thus, our possibility horizon should include the following four worlds, characterised by their complete states at time t_1:

	Complete state at time t_1:	*Events at* t_1:	t_4:
@:	**A** fires, **B** fires, **C** fires	A, B, C	I
w_1:	**A** fires, **B** does not fire, **C** fires	A, B, C	I
w_2:	**A** fires, **B** fires, **C** does not fire	A, B, C	I
w_3:	**A** fires, **B** does not fire, **C** does not fire	A, B, C	I

Call this possibility horizon $\mathscr{H}$. We may represent the relations of comparative distance-at-t_1 between these worlds as follows, where the worlds in which I does not occur are coloured grey:

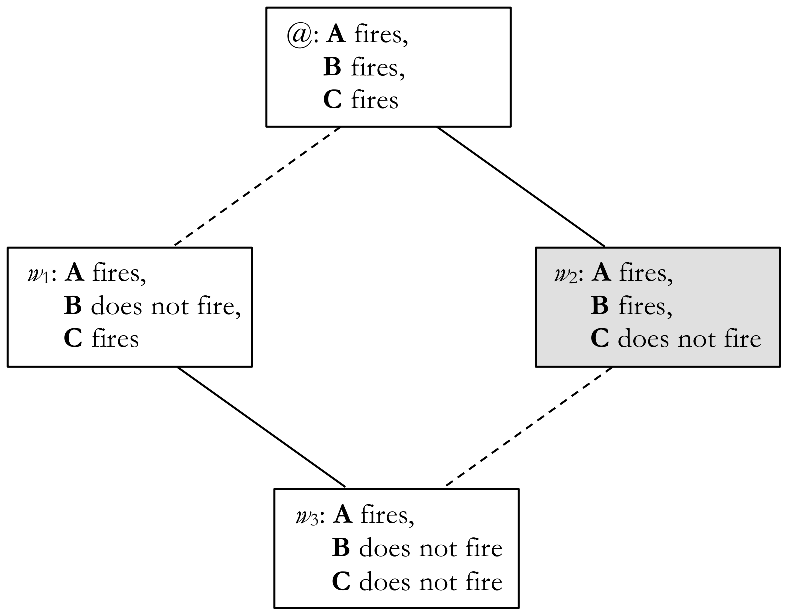


Figure 8

To see that I does not security-depend on B within $\mathscr{H}$, we now proceed in the usual three steps:

Step 1: our first step is to evaluate I's security at t_1 in @. From the standpoint of @, there is one minimal dependence set for I at t_1 – namely, $\{C\}$; and the closest-at-t_1 world where C fails to occur is w_2. Thus, I's security at t_1 in @ is given by the distance-at-t_1 from @ to w_2.

Step 2: our second step is to evaluate I's security at t_1 in the closest-at-t_1 world(s) in which B does not occur. From the standpoint of @, the closest-at-t_1 world where B does not occur is w_1. From the standpoint of w_1, there is one minimal dependence set for I at t_1 – namely, $\{\neg B, C\}$; and the closest-at-t_1 world in which $\neg B$ and C both fail to occur is w_2. Thus, I's security at t_1 in w_1 is given by the distance-at-t_1 from w_1 to w_2.

Step 3: our final step is to compare I's security at t_1 in @ with I's security at t_1 in w_1. As indicated in the figure, the distance-at-t_1 from @ to w_2 is strictly shorter than the distance-at-t_1 from w_1 to w_2. Thus, we find that I is *more secure* at t_1 in w_1 than it is in @.

This yields the result that *I* does not security-depend on *B* within $\mathscr{H}$.[11] And for this reason, *B* is not a cause of *I*, since the necessary condition of security-dependence fails to be satisfied.

2.4 Two kinds of counterexamples to transitivity

On my proposed account of causation, the two conditions of process-connection and security-dependence are individually necessary and jointly sufficient for causation. This commits me to the following generalisation: in *all* counterexamples to *Transitivity of causation*, causation fails to behave transitively because security-dependence fails to behave transitively. In all these cases, then, the non-transitivity of security-dependence allows us to understand why causation fails to behave transitively.

If this is correct, it holds an additional lesson. As I mentioned above, counterexamples to *Transitivity of causation* are standardly divided into two groups: counterexamples that do not involve omissions, and counterexamples that do.[12] In keeping with my general suggestion that there is no deep metaphysical distinction between so-called positive events and omissions, this distinction plays no role in my proposed treatment of the counterexamples. Instead, my treatment of these cases draws attention to another way of categorising the counterexamples to *Transitivity of causation*: based on *how* the effect fails to security-depend on the candidate cause.

There are two ways in which a later event *e* may fail to security-depend on an earlier event *c*, when the relevant possibility horizon includes a world where *c* does not occur: it may be the case that *e* is security-independent from *c*, or it may be the case that *e* security-depends negatively on *c*. In cases of security-independence, *c* is causally *irrelevant* to *e*; in cases of negative security-

---

[11] Within the larger possibility horizon where *A* is also treated as a candidate cause, we similarly find that *I* security-depends negatively on *B*. Within the smaller possibility horizon where *C* is treated as a background condition, we find that *I* is security-independent from *B* (this corresponds to the variation of *Boulder* discussed in section 2.4 below). In both cases, the condition of security-dependence fails to be satisfied.
[12] Hall (2000), Paul and Hall (2013), pp. 215-37.

dependence, *c* has negative causal relevance to *e* and, correspondingly, it seems intuitively correct to say that *e* occurs *in spite of c* (cf. Chapter 8 section 3.2).[13]

The two cases we have considered above illustrate the two ways in which one event can fail to security-depend on another: in *Switch*, we found that *F* would have been *just as secure* at t_1 if *B* had not occurred. In *Switch*, then, *B* is causally *irrelevant* to *F*. This fits nicely with our intuitive judgement: intuitively, we want to say that *B* (Suzy's flipping the switch) is entirely irrelevant to *F* (the train's arrival). In *Boulder*, on the other hand, we found that *I* would have been *more secure* at t_1 if *B* had not occurred. In this case, then, *B* has negative causal relevance to *I*. Once again, this fits nicely with our intuitive judgement: intuitively, we want to say that *I* (Hiker's survival) occurs *in spite of B* (the boulder's fall).

On this basis, we may divide counterexamples to *Transitivity of causation* into two groups: i) cases where transitivity fails because the candidate cause is causally irrelevant to the effect, and ii) cases where transitivity fails because the candidate cause has negative causal relevance to the effect.

It is important to note that this distinction cuts across the distinction between counterexamples to *Transitivity of causation* based on whether or not they involve omissions. Above, we have considered one case that does not involve omissions, namely *Switch*, where Suzy's flipping the switch is simply irrelevant to the train's arrival. And we have considered one omission-involving case, namely *Boulder*, where Hiker survives in spite of the boulder's fall. However, we may easily tinker with these cases so as to get a version of *Switch* where the train arrives in spite of Suzy's flipping the switch, and a

[13] In addition, it may also be the case that the possibility horizon under consideration contains two or more closest-at-*t* not-*c*-worlds, where these worlds do not agree on whether *e* is *less secure*, *just as secure* or *more secure* at *t*. Such cases require a contrastive treatment, where we restrict the possibility horizon under consideration until it yields uniform results (cf. Northcott (2008), p. 120). If *e* still fails to security-depend on *c* within a restricted possibility horizon $\mathscr{H}$ that yields uniform results, this is either because *e* is security-independent from *c* within $\mathscr{H}$ or because *e* security-depends negatively on *c* within $\mathscr{H}$. For more on cases that require a contrastive treatment, see Chapter 10 section 3.

version of *Boulder* where the boulder's fall is simply irrelevant to Hiker's survival.

We may get a version of *Switch* where the train arrives *in spite of* Suzy's flipping the switch by making the following addition:

> *Addition to Switch*: When the train is approaching, the left track is in fact disconnected. Just as Suzy flips the switch, Sally comes by and reconnects the track. If Sally had not reconnected the track, the train would have been derailed.[14]

The structure of this case is represented below:

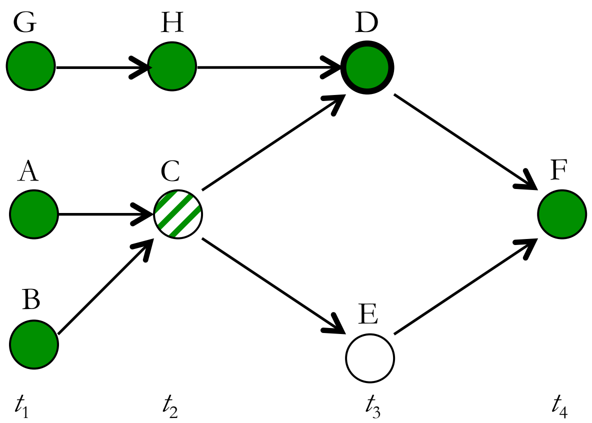


Figure 9

G here represents Sally's reconnecting the tracks. Furthermore, **D** is a stubborn neuron representing the fact that the train's journey down the left-hand track occurs just in case Suzy flips the switch *and* Sally reconnects the tracks.

The relevant possibility horizon $\mathscr{H}$ for determining whether *F* security-depends on *B* includes the following eight worlds:

[14] This case is based on a case presented by Sartorio. Comparing this case with *Switch*, Sartorio notes that '[i]f anything, we feel even more reluctant to say that [Suzy's] flipping the switch is a cause […] in this case' (Sartorio (2005), p. 82). This comment supports my claim that the train here arrives *in spite of* Suzy's flipping the switch.

	Complete state at time t_1:	*Events at t_1:*	t_4:
@:	**A** fires, **B** fires, **G** fires	A, B, G	F
w_1:	**A** does not fire, **B** fires, **G** fires	A, B, G	F
w_2:	**A** fires, **B** does not fire, **G** fires	A, B, G	F
w_3:	**A** does not fire, **B** does not fire, **G** fires	A, B, G	F
w_4:	**A** fires, **B** fires, **G** does not fire	A, B, G	F
w_5:	**A** does not fire, **B** fires, **G** does not fire	A, B, G	F
w_6:	**A** fires, **B** does not fire, **G** does not fire	A, B, G	F
w_7:	**A** does not fire, **B** does not fire, **G** does not fire	A, B, G	F

We may represent the comparative distance-at-t_1 between these eight worlds as follows (note that worlds where F does not occur are coloured grey):

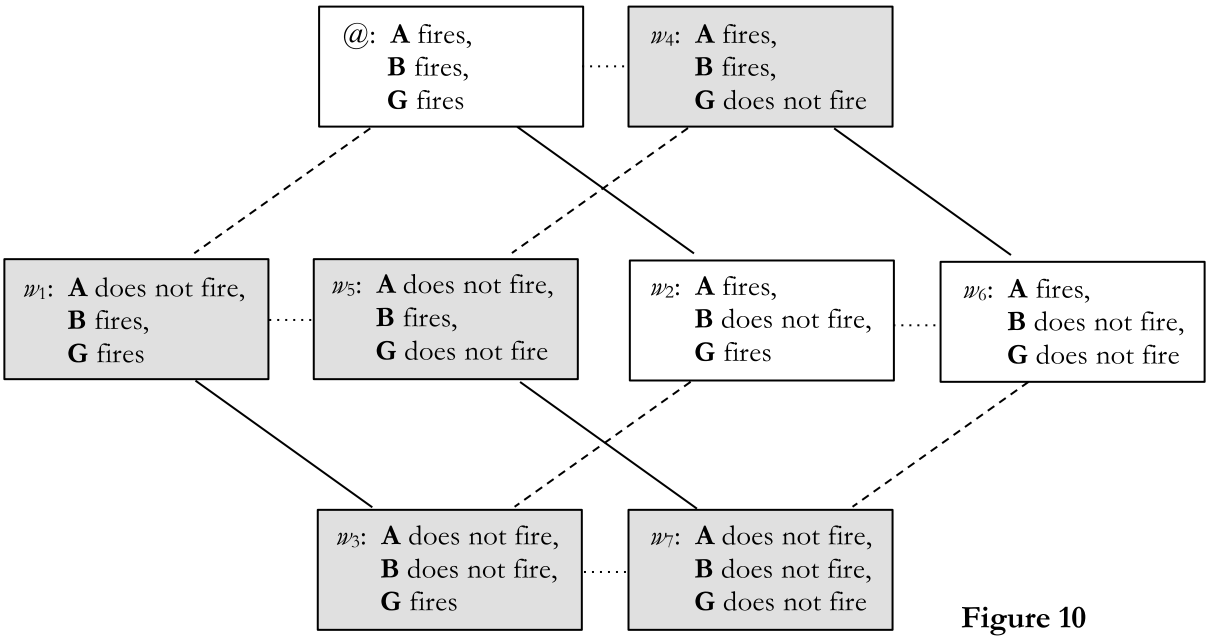


Figure 10

Following our usual three-step procedure, we now find that F (the train's arrival) security-depends negatively on B (Suzy's flipping the switch) within $\mathscr{H}$: if B had not occurred, F would have been *more secure* at t_1.

Step 1: our first step is to evaluate F's security at t_1 in @. From the standpoint of @, there are two minimal dependence sets for F at t_1 – namely,

$\{A\}$ and $\{G\}$. The closest-at-t_1 world where A fails to occur is w_1, and the closest-at-t_1 world where G fails to occur is w_4. Thus, F's security at t_1 in @ is given by the distance-at-t_1 from @ to w_1 together with the distance-at-t_1 from @ to w_4.

Step 2: our second step is to evaluate F's security at t_1 in the closest-at-t_1 world(s) where B does not occur. From the standpoint of @, the closest-at-t_1 world where B does not occur is w_2. From the standpoint of w_2, F has two minimal dependence sets at t_1 – namely, $\{A\}$ and $\{\neg B, G\}$; the closest-at-t_1 world where A fails to occur is w_3, and the closest-at-t_1 world where both $\neg B$, and G fail to occur is w_4. Thus, F's security at t_1 in w_2 is given by the distance-at-t_1 from w_2 to w_3 together with the distance-at-t_1 from w_2 to w_4.

Step 3: our third and final step is to compare F's security at t_1 in @ with F's security at t_1 in w_2. As indicated in the figure, the distance-at-t_1 from @ to w_1 is exactly the same as the distance-at-t_1 from w_2 to w_3. Furthermore, the distanc-at-t_1 from @ to w_4 is strictly shorter than the distance-at-t_1 from w_2 to w_4. From this, it follows that F is *more secure* at t_1 in w_2 than it is at t_1 in @.

In the present case, F thus security-depends negatively on B – fitting nicely with the intuitive verdict that F (the train's arrival) here occurs *in spite of* B (Suzy's flipping the switch).[15]

We may similarly tinker with *Boulder* to get a version of the case where the boulder's fall is simply irrelevant to Hiker's survival: to get this, we simply need to ensure that Hiker will *automatically* duck if the boulder falls, irrespective of the circumstances. Setting up the case in this way, we get the structure illustrated below:[16]

[15] Within the smaller possibility horizon where G is treated as a background condition, the case behaves like the original version of *Switch*, and we here find that F is security-independent from B.

[16] Cf. Paul and Hall (2013), Figure 10.

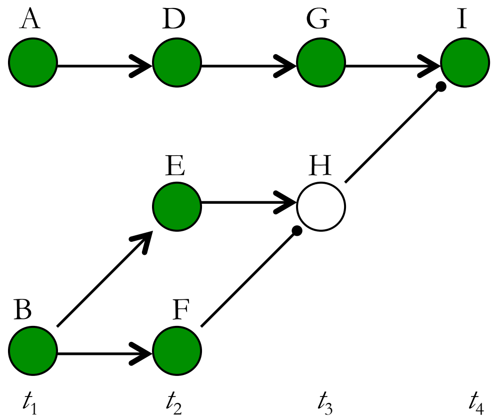


Figure 11

In this case, the relevant possibility horizon $\mathscr{H}$ for determining whether *I* security-depends on *B* includes just two worlds, namely:

	Complete state at time t_1:	*Events at* t_1:	t_4:
@:	**A** fires, **B** fires	*A*, *B*	*I*
w_2:	**A** fires, **B** does not fire	*A*, *B*	*I*

Since *I* occurs in both of these worlds, we find that *I* is maximally secure at t_1 in both @ and w_2. And thus we find that *I* would have been *just as secure* at t_1 if *B* had not occurred, fitting nicely with the intuitive judgement that *B* (the boulder's fall) is here simply *irrelevant* to *I* (Hiker's survival).[17]

Together, these two cases show that my classification of transitivity-failure cases based on *how* the effect fails to security-depend on the candidate cause *cuts across* the distinction between cases that do and do not involve omissions. And while it makes no difference to my proposed way of handling these cases whether or not they involve omissions, it does make a difference *how* the effect fails to security-depend on the candidate cause: the distinction between the two ways in which the effect may fail to security-depend on its candidate cause captures the distinction between cases where the candidate cause is causally irrelevant and cases where it has negative causal relevance.

[17] When we consider a larger possibility horizon, which includes worlds where *A* does not occur, we similarly find that *I* is security-independent of *B*.

2.5 Borderline cases

The structure of early preemption cases is remarkably similar to the structure of transitivity failure cases. To see this similarity, consider our standard case of early preemption:

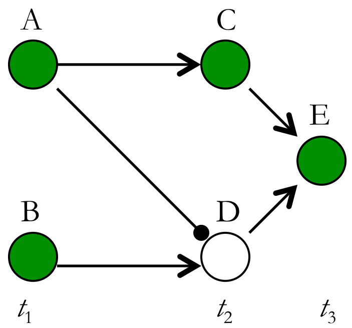


Figure 12

In standard preemption cases, a candidate cause *c* is connected to an effect *e* by a chain of counterfactual dependence, even though *e* does not depend counterfactually on *c*. In the case above, for example, *E* depends counterfactually on *C*, and *C* depends counterfactually on *A*, but *E* does not depend counterfactually on *A*. *Switch* and *Boulder* exhibit just these structural features as well.

In light of these structural similarities, we may think of the condition of security-dependence as a way of *drawing a line* separating early preemption cases from transitivity failure cases: what separates our case of early preemption from the counterexamples to transitivity is that the effect security-depends on the candidate cause in the case of early preemption, while the effect fails to security-depend on the candidate cause in the counterexamples to transitivity.

Once we recognise the similarities between the two kinds of cases, however, we should also expect to find borderline cases: cases we may easily classify as curious variants of early preemption, even though they are in fact transitivity failure cases, or *vice versa*. In the following, I give an example of such a case, and two more examples of such cases may be found in Appendix B (case 23 and 24):

> *Window-shattering*: Suzy throws a rock at the window. Her aim is slightly off, but a strong gust of wind brings her rock back on course. It hits the window and the window shatters. If Suzy had not thrown, Billy would have thrown a larger rock and hit the window, independently of whether there was wind or not.[18]

The structure of the case is represented below:

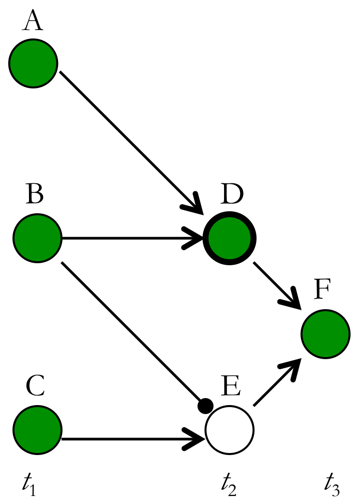


Figure 13

A here represents the strong gust of wind, *B* represents Suzy's throw, *C* represents Billy's readiness to throw, and *D* represents Suzy's rock being on target to hit the window. Note that **D** is a stubborn neuron that requires two stimulatory signals in order to fire.

The present case has the structural features shared by preemption cases and transitivity failure cases: *F* depends counterfactually on *D*, and *D* depends counterfactually on *B*, but *F* does not depend counterfactually on *B*. We may now ask: is *B* a cause of *F*? It seems to me that we are to some extent ambivalent about how to answer this question. I take it to be an advantage of my condition of security-dependence that it can capture this ambivalence, as we shall now see.

---

[18] Note that this case is similar to cases of alleged chance-lowering causation. For discussion, see e.g. Dowe (2000), pp. 33-40; Hitchcock (2004); Lewis (1986*e*), pp. 179-80; Mellor (1995), pp. 67-68.

First, note that B is process-connected to F via the following process:

t_1:	B	(together with A)
t_2:	D	
t_3:	F	

This process is illustrated below:

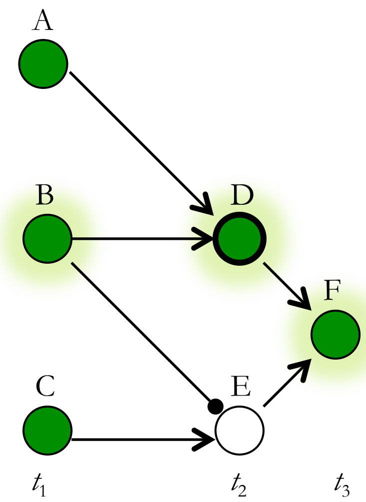


Figure 13*

Thus, the condition of process-connection cannot on its own capture why we hesitate about whether to count B as a cause of F. To capture this, we need the condition of security-dependence. To see this, let us consider the most natural choice of possibility horizon, namely the possibility horizon $\mathscr{H}$ containing the following eight worlds:

	Complete state at time t_1:	*Events at* t_1:	t_3:
@:	**A** fires, **B** fires, **C** fires	A, B, C	F
w_1:	**A** does not fire, **B** fires, **C** fires	A, B, C	F
w_2:	**A** fires, **B** does not fire, **C** fires	A, B, C	F
w_3:	**A** does not fire, **B** does not fire, **C** fires	A, B, C	F
w_4:	**A** fires, **B** fires, **C** does not fire	A, B, C	F
w_5:	**A** does not fire, **B** fires, **C** does not fire	A, B, C	F

w_6:	**A** fires, **B** does not fire, **C** does not fire	*A*, *B*, *C*	*F*
w_7:	**A** does not fire, **B** does not fire, **C** does not fire	*A*, *B*, *C*	*F*

We may illustrate the distances-at-t_1 between these eight worlds as follows:

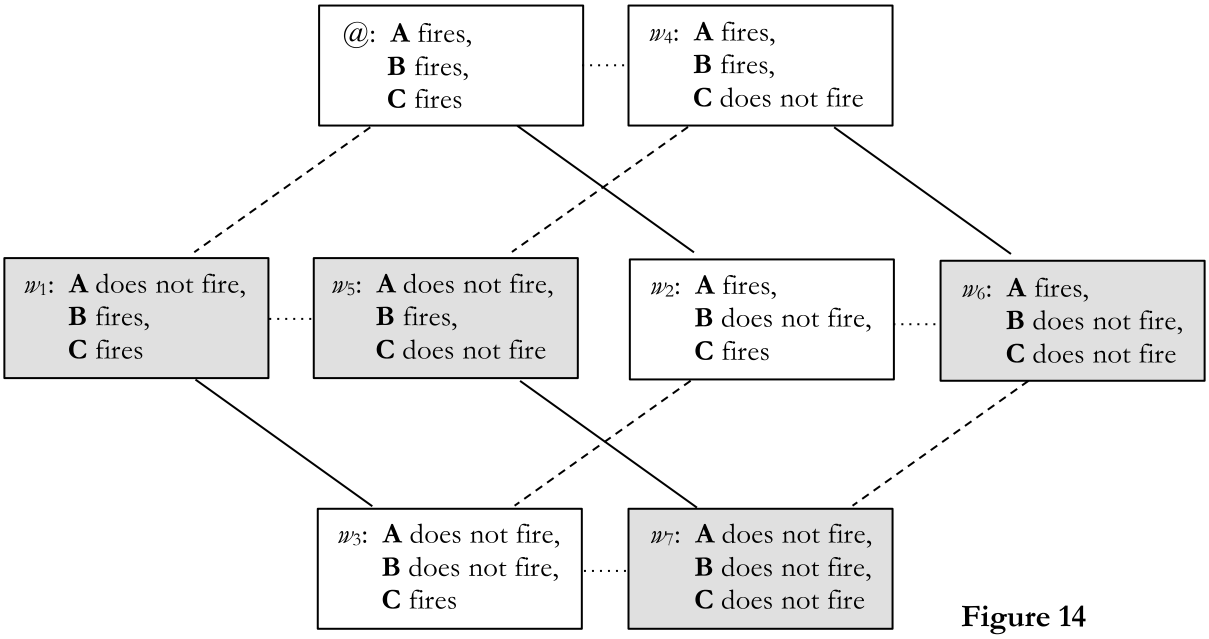


Figure 14

Call this possibility horizon $\mathscr{H}$. Within this possibility horizon, we now find that it comes down to a subtle judgement of comparative security whether or not *F* security-depends on *B* within $\mathscr{H}$. To bring this out, let us go through our usual three steps.

Step 1: the first step is to evaluate *F*'s security at t_1 in @. From the standpoint of @, there are two minimal dependence sets for *F* at t_1 – namely, {*A*} and {*B*, *C*}. The closest-at-t_1 world where *A* does not occur is w_1, and the closest-at-t_1 world where both *B* and *C* do not occur is w_6. Thus, *F*'s security at t_1 in @ is given by the distance-at-t_1 between @ and w_1 together with the distance-at-t_1 between @ and w_6.

Step 2: the second step is to evaluate *F*'s security at t_1 in the closest-at-t_1 world(s) where *B* does not occur. From the standpoint of @, the closest-at-t_1

world where B does not occur is w_2. From the standpoint of w_2, F has one minimal dependence set at t_1 – namely, $\{C\}$; and the closest-at-t_1 world where C does not occur is w_6. Thus, F's security at t_1 in w_2 is given by the distance-at-t_1 between w_2 and w_6.

Step 3: the third step is to compare F's security at t_1 in @ with F's security at t_1 in w_2. We here find that it depends on very specific details of the case how we should make this comparison.

If there is a sufficiently long distance-at-t_1 from @ to w_1, which is the closest-at-t_1 world where A (the strong gust of wind) does not occur, then we find that F security-depends on B – and if so, the case should be classified as a curious variant of early preemption. If, on the other hand, the distance-at-t_1 from @ to w_1 is relatively short, we find that F is *just as secure* or *more secure* at t_1 in w_2. This implies that B is not a cause of F – and the case should be classified as a transitivity failure case. I believe that these results correctly reflect our intuitive judgements about the case.[19]

2.6 Does security-dependence capture all transitivity failure cases?

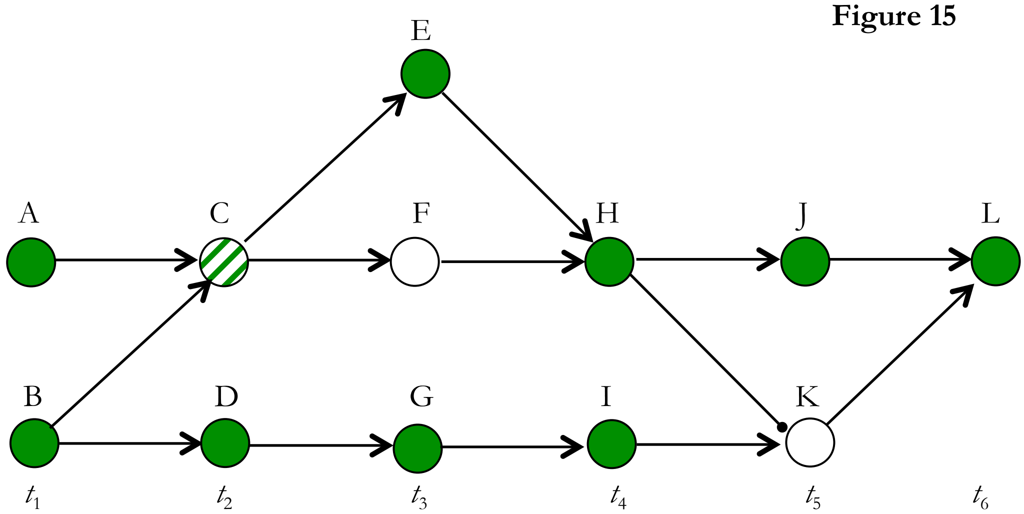
So far, we have seen a number of cases where causation fails to behave transitively, and where this is explained by the fact that security-dependence

[19] Even when the situation is such that the strong gust of wind could very easily have been absent, we may still judge that there is a sense in which Suzy's throw is a cause of the window-shattering. My account of causation can capture this judgement as being about a *restricted* possibility horizon: the strong gust of wind (**A**'s firing) is clearly a deviant event. Our contextually determined possibility horizon should therefore include worlds where there is no gust of wind (**A** does not fire). However, we may wish to draw attention to the fact that while Suzy's throw puts the window-shattering at risk in one way – since it makes it dependent on the strong gust of wind – there is another threat that Suzy's throw makes more remote, namely the threat that *no one* throws a rock at the window. To capture this, we may artificially restrict our possibility horizon to only include worlds where the strong gust of wind occurs (where **A** fires). In fact, we have a ready-made linguistic mechanism for making this restriction: by using phrases such as 'given that . . .', as in 'given that the strong gust of wind was there, Suzy's throw caused the window-shattering'. Within such a restricted possibility horizon, where the strong gust of wind occurs in all worlds (A occurs in all worlds), we straightforwardly find that the window-shattering (F) security-depends on Suzy's throw (B). Thus, the hedged claims – 'given that the strong gust of wind was there, Suzy's throw caused the window-shattering', and 'given that A occurred, B caused F' – come out true.

fails to behave transitively. However, my suggestion that process-connection and security-dependence are individually necessary and jointly sufficient for causation commits me to a bolder claim: namely, that *all* cases where causation fails to behave transitively are cases where security-dependence fails to behave transitively.

By building together transitivity-failure cases and preemption cases, we can arrive at apparent counterexamples to this claim.[20] Consider, for example, the figure below:

Figure 15



The structure of neurons comprising **A**, **B**, **C**, **E**, **F**, and **H** corresponds to one of our standard transitivity-failure cases, namely *Switch*. Furthermore, the structure of neurons comprising **H**, **I**, **J**, **K**, and **L** corresponds to a standard case of early preemption. Combining our intuitive judgements about these two cases, it seems intuitively plausible to hold that *B* is *not* a cause of *L*, although *B* causes *E*, *E* causes *H*, and *H* causes *L*.

As is easily verified, however, *B* satisfies the two conditions of process-connection and security-dependence: *B* is process-connected to *L*, and *L* security-depends on *B* within the contextually relevant possibility horizon. Thus, the case presents an apparent counterexample my proposed account of

[20] I am grateful to Derek Ball for making me aware of such cases.

causation and, in particular, to my claim that my condition of security-dependence can account for all transitivity-failure cases. My response has two parts:

First, it is not so clear that we are able to simultaneously pay attention to all the relevant features of the case when we reach our intuitive judgement that *B* is not a cause of *L*. It is obviously true that when a candidate cause *merely* functions as a switch, then it is not a cause – since it makes no difference. If we focus on the features of Figure 15 where *B* merely functions as a switch, we therefore reach the judgement that *B* is not a cause of *L*. Similarly, it is obviously true that when a candidate cause *merely* functions as a preempted backup, it is not a cause – since it is not part of a process leading up to the effect. If we focus on the features of Figure 15 where *B* merely functions as a preempted backup, we therefore also reach the judgement that *B* is not a cause of *L*. But in reaching both of these judgements, we focus on certain features of the case, while neglecting others. And it is not clear to me that we are able to reach a firm and direct intuitive judgement that is based on an appreciation of the case as a whole. This suggests that we should not put too much weight on our intuitive judgement on this case.

Secondly, it is important to keep in mind that my project is to find the candidate meaning of 'cause' that strikes the best balance between eligibility and charity to use. Given this project, intuitions about specific cases sometimes have to be overridden, if accommodating them would be too costly. I believe this is the case here: as we shall see in Chapter 11, my claim that security-dependence can account for *all* transitivity failure cases supports a restricted principle of transitivity, which legitimises all or almost all of our ordinary practices of appealing to transitivity in our causal reasoning. Accommodating our intuition in the present case would force us to give up on that restricted principle of transitivity. I do not believe that would be worth the cost.

3. The non-intrinsicness of causation

There is a close connection between counterexamples to the transitivity of causation and counterexamples to the intrinsicness of causation. Indeed, slight modifications of the very same case – namely, *Switch* – provide counterexamples to both transitivity and intrinsicness. For this reason, it should perhaps not come as a surprise that the condition of security-dependence can also accommodate our intuitive verdicts in counterexamples to the intrinsicness of causation. My aim in this section is to demonstrate this result.

I begin with a general explanation of how causation fails to be intrinsic to a process (section 3.1). I then consider two counterexamples to the intrinsicness of causation: in parallel with counterexamples to transitivity, counterexamples to intrinsicness may be divided into two groups – counterexamples that do not involve omissions, and counterexamples that do. A small modification of *Switch* provides a counterexample to intrinsicness that does not involve omissions. And our standard case of double prevention provides an omission-involving counterexample to intrinsicness (considered in Chapter 2 section 1.3.3). In section 3.2 and 3.3, I show how the condition of security-dependence yields intuitively correct verdicts on both of these cases, irrespective of whether they involve omissions or not.

3.1 How intrinsicness fails

The intuitive principle that causation is intrinsic to a process may be given a preliminary statement as follows (the asterisk indicates that I do not endorse this principle):

> **Intrinsicness of causation*: if *c* causes *e*, and a structure of events $\mathscr{S}$, including all the events that are involved in a process connecting *c* to *e*, is governed by the same laws and exactly matches a structure of events $\mathscr{S}$*, then the counterpart *c** of *c* in $\mathscr{S}$* is a cause of the counterpart *e** of *e* in $\mathscr{S}$*.

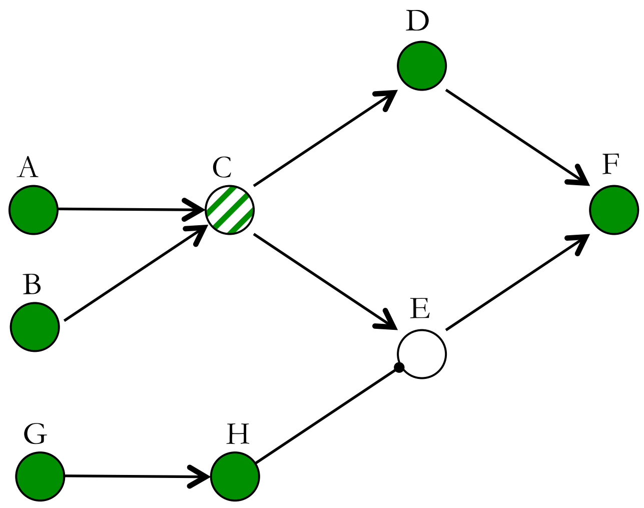
As we have already seen, there are counterexamples to *Intrinsicness of causation* on any reasonable way of spelling out exactly which events are involved in the process from one event to another. The condition of process-connection cannot on its own accommodate or explain our intuitive verdicts on these counterexamples. For, as I show in Appendix A section 3, the relation of process-connection is itself intrinsic to a process (on this, see also Chapter 11 section 3).

To accommodate and explain our intuitive judgements, we therefore need the condition of security-dependence: as we shall see in the following, the relation of security-dependence is *not* intrinsic to a process. Let us say that causation fails to behave intrinsically when *c* is a cause of *e*, and the process connecting *c* to *e* is just like the process connecting *c** to *e**, but *c** does not cause *e**. And let us say that security-dependence fails to behave intrinsically when *e* security-depends on *c*, and the process connecting *c* to *e* is just like the process connecting *c** to *e**, but *e** does not security-depend on *c**. Then my suggestion is that causation fails to behave intrinsically in just those cases where security-dependence fails to behave intrinsically, and the failure of security-dependence to behave intrinsically explains why causation fails to behave intrinsically in these cases.[21]

3.2 Switch

Let us begin by considering *Switch* once again – this time, as a counterexample to *Intrinsicness of causation*. Consider first the following modified version of *Switch*: everything is as in the original case, except that there is a fallen tree lying across the right-hand track, such that the train would have been derailed if Suzy had not flipped the switch. The structure of this case is illustrated below, with *G* representing the presence of the fallen tree:

[21] Note that security-dependence may fail to behave intrinsically for two reasons: i) merely because the change in setting means that a different world within a given possibility horizon is the actual world, or ii) because the change in setting induces a shift in which worlds are included in the contextually relevant possibility horizon.

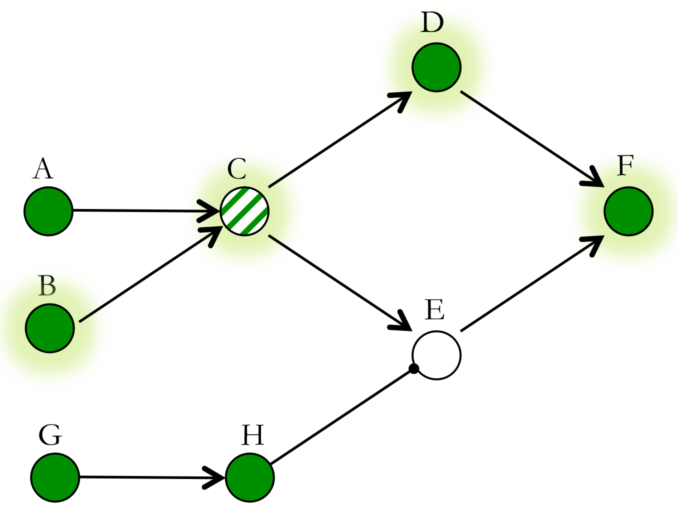


t_1 t_2 t_3 t_4 **Figure 16**

Intuitively, *B* is here a cause of *F*. Correspondingly, we find that *B* is process-connected to *F* via the following process:

t_1:	*B*	(together with *A*)
t_2:	*C-stripes*	
t_3:	*D*	
t_4:	*F*	

This is illustrated below:



t_1 t_2 t_3 t_4 **Figure 16***

On any reasonable way of spelling out which events are involved in the process from *B* to *F*, it is surely the case that *G* (the presence of the fallen tree) is *not* involved. Removing the fallen tree – letting **G** be dormant instead of firing – therefore counts as a change that is entirely extrinsic to the process connecting *B* to *F*. But by carrying out this change, we now get back to a case that is essentially equivalent to the version of *Switch* we have already considered: for in this case, there is no fallen tree on the right-hand track, and so the two tracks are equally good – just as they are in the original case.

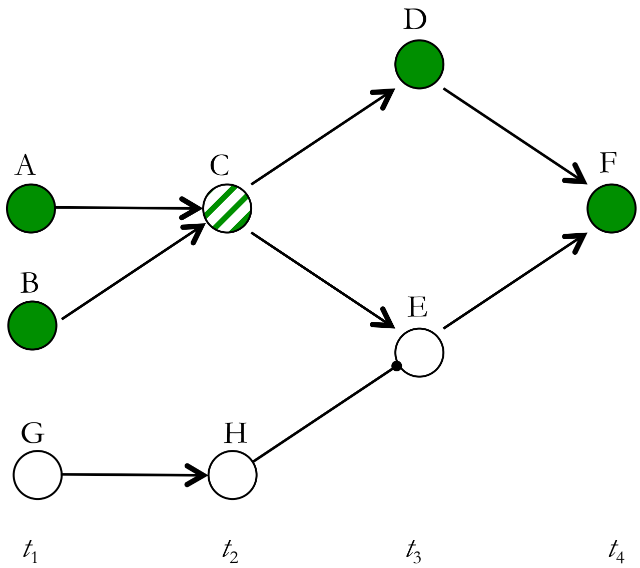


Figure 17

Together, these two cases present a counterexample to *Intrinsicness of causation*: *B* causes *F* in Figure 16, and the structure of events $\mathcal{S}$ = {*A*, *B*, *C-stripes*, *D*, *F*} in Figure 16 includes all the events involved in the process from *B* to *F*, and is governed by the same laws and exactly matches the structure of events $\mathcal{S}$* = {*A*, *B*, *C-stripes*, *D*, *F*} in Figure 17. By intrinsicness, it therefore follows that *B* is a cause of *F* in Figure 17. But intuitively, this is false: *B* is *not* a cause of *F* in Figure 17 – rather, *B* is simply *irrelevant* to *F*.

The condition of process-connection cannot on its own accommodate this intuitive verdict or explain why causation fails to be intrinsic in this case, since process-connection is itself intrinsic to a process (cf. Appendix A section 3). Thus, we find that *B* is also process-connected to *F* in Figure 17 – as indeed it is, via the process:

t_1:	*B*	(together with *A*)
t_2:	*C-stripes*	
t_3:	*D*	
t_4:	*F*	

To get the intuitively correct verdict, we therefore need the further condition of security-dependence. The condition of security-dependence allows us to accommodate our intuitive judgement for the simple reason that security-dependence is *not* intrinsic to a process: in Figure 16, *F* depends counterfactually on *B* within any possibility horizon that treats *B* as a candidate cause, and from this it immediately follows that *F* security-depends on *B* within any such possibility horizon. In Figure 17, by contrast, we find that *F* does *not* security-depend on *B* within the contextually relevant possibility horizon – rather, *F* would have been *just as secure* at t_1 if *B* had not occurred.

To see this, note that it is natural to treat $\neg G$, i.e. the absence of a fallen tree, as a background condition. Thus, the contextually relevant possibility horizon includes just the following four possible worlds:

	Complete state at time t_1:	*Events at t_1:*	t_4:
@:	**A** fires, **B** fires, **G** does not fire	A, B, $\neg G$	F
w_1:	**A** does not fire, **B** fires, **G** does not fire	A, B, $\neg G$	F
w_2:	**A** fires, **B** does not fire, **G** does not fire	A, B, $\neg G$	F
w_3:	**A** does not fire, **B** does not fire, **G** does not fire	A, B, $\neg G$	F

We may represent the distance-at-t_1 between these worlds as follows:

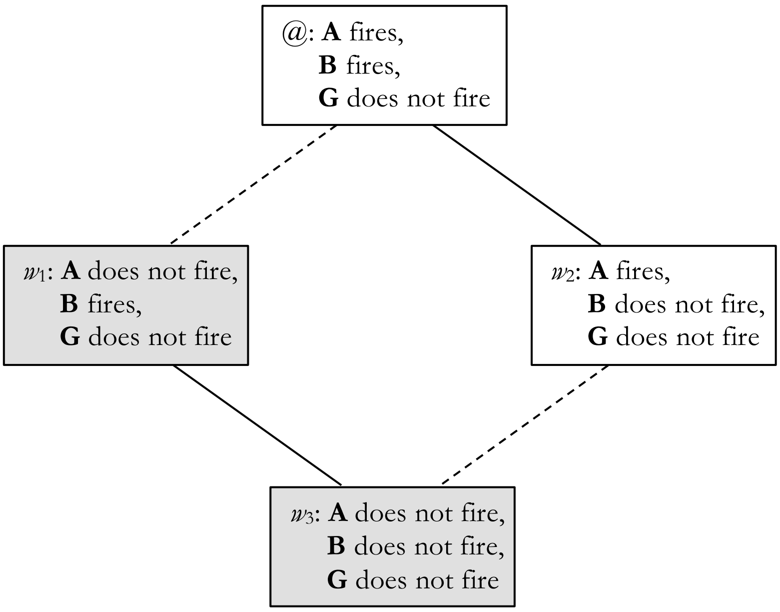


Figure 18

By exactly the same arguments as in the original version of *Switch*, we now find that *F* would have been *just as secure* at t_1 if *B* had not occurred.[22]

In this way, the condition of security-dependence allows us to maintain the intuitively correct verdict: *B* is a cause of *F* in Figure 16, since *B* here satisfies the two necessary and jointly sufficient conditions of process-connection and security-dependence. By contrast, *B* is *not* a cause of *F* in Figure 17: for although *B* is still process-connected to *F*, *F* does not security-depend on *B*.

[22] Note that this result is dependent on our particular choice of possibility horizon: within a larger possibility horizon, which includes worlds where **G** fires, *F* does security-depend on *B*. More generally, then, the difference between the case where **G** fires (Figure 16) and the case where **G** does not fire (Figure 17) is the following: in the case where **G** fires, *any* possibility horizon that treats *B* as a candidate cause yields the result that *F* security-depends on *B*. By contrast, in the case where **G** does not fire, the verdict on whether *F* security-depends on *B* is much more sensitive to our choice of possibility horizon: within the larger possibility horizon, which includes worlds where **G** fires, *F* still security-depends on *B*; but within the smaller and more natural choice of possibility horizon, where **G**'s failure to fire is treated as a background condition, we find that *B* is simply irrelevant to *F*.

3.3 Double prevention

Next, let us consider an omission-involving counterexample to *Intrinsicness of causation*. In particular, let us consider the following case of double prevention (discussed in Chapter 2 section 1.3.3):

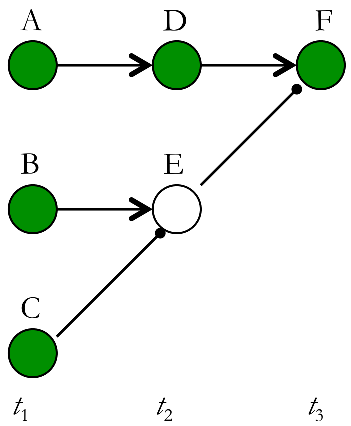


Figure 19

It here seems intuitively correct to say that C is a cause of F. Correspondingly, we find that C is process-connected to F via the following process:

t_1: C

t_2: $\neg E$ (together with D)

t_3: F

This is illustrated below:

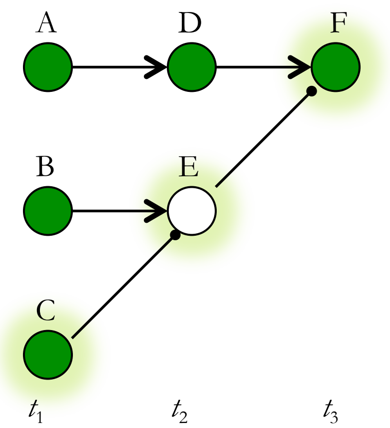


Figure 19*

On any reasonable way of spelling out what it takes for an event to be involved in the process connecting *C* to *F*, this process does not involve *B*. Thus, letting **B** be dormant instead of firing is a change that is entirely extrinsic to the process connecting *C* to *F*. This now yields the following case:

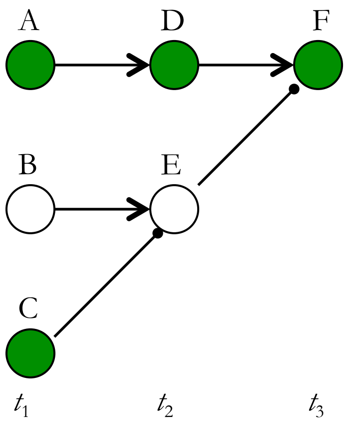


Figure 20

Together, these two cases provide a counterexample to *Intrinsicness of causation*: *C* is a cause of *F* in Figure 19, and the structure of events $\mathscr{S} = \{C, D, \neg E, F\}$ in Figure 19 includes all the events involved in the process from *C* to *F*, and is governed by the same laws and exactly matches the structure of events $\mathscr{S}^* = \{C, D, \neg E, F\}$ in Figure 20. By intrinsicness, it therefore follows that *C* is a cause of *F* in Figure 20. Intuitively, however, this is false: as Paul and Hall note, letting **B** be dormant 'completely reverses the intuitive verdict about *C* – once we remove the threat created by **B**'s firing, we intuitively judge that *C* is completely idle with respect to [*F*]'.[23]

The condition of process-connection cannot on its own secure this verdict. For, as mentioned above, the relation of process-connection is itself intrinsic to a process. Thus, we find that *C* is also process-connected to *F* in Figure 20, via the following process:

[23] Paul and Hall (2013), p. 197. Paul and Hall are here commenting on a version of the case where neuron **B** is not merely dormant, but has been removed entirely. However, their reasoning applies equally well to the case presented in Figure 20, since letting **B** be dormant is enough to remove the threat created by **B**'s firing.

t_1: C
t_2: $\neg E$ (together with D)
t_3: F

This process is illustrated below:

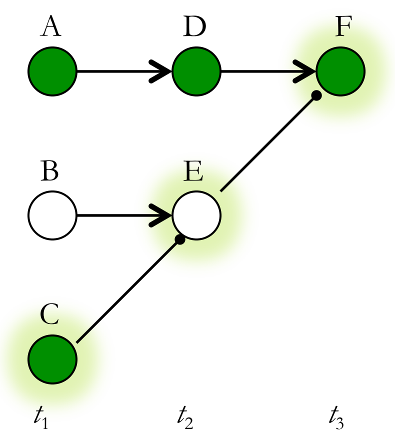


Figure 20*

As this shows, the condition of process-connection cannot on its own yield the intuitively correct verdict on this case. We need the further condition of security-dependence: as I will now show, F security-depends on C within the contextually relevant possibility horizon in Figure 19, whereas F does *not* security-depend on C within the contextually relevant possibility horizon in Figure 20.

It is easy to see that F security-depends on C in Figure 19: within any possibility horizon that treats C as a candidate cause, F depends counterfactually on C in Figure 19, and it follows immediately from this that F security-depends on C. In the case of Figure 20, it is natural to treat $\neg B$ as a background condition. This yields the following possibility horizon $\mathscr{H}$:

	Complete state at time t_1:	*Events at t_1:*	t_3:
@:	**A** fires, **B** does not fire, **C** fires	A, $\neg B$, C	F
w_1:	**A** does not fire, **B** does not fire, **C** fires	A, $\neg B$, C	F
w_2:	**A** fires, **B** does not fire, **C** does not fire	A, $\neg B$, C	F

w_3: **A** does not fire, **B** does not fire, **C** does not fire A, $\neg B$, C F

Call this possibility horizon $\mathscr{H}$. We may represent the comparative distances-at-t_1 between these worlds as follows:

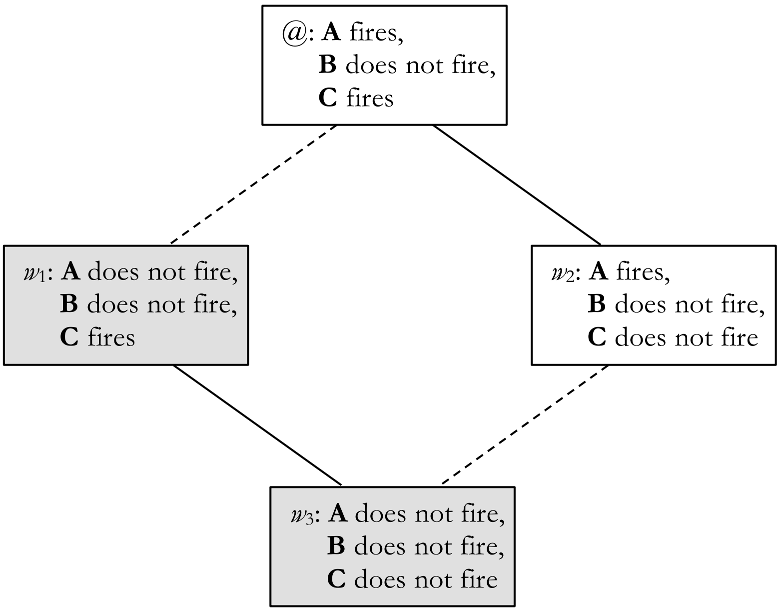


Figure 21

By exactly the same arguments as in *Switch*, we now find that F would have been *just as secure* at t_1 if C had not occurred. Thus, F is security-independent of C within possibility horizon $\mathscr{H}$. This result fits nicely with the nuances of our intuitive judgement: it seems intuitively correct to say that C is causally *irrelevant* to F – as Paul and Hall write, 'C is completely idle with respect to [F]'[24] – and this is exactly what we have found.[25]

---

24 Paul and Hall (2013), p. 197.

25 Note that this result depends on our particular choice of possibility horizon: within a larger possibility horizon, which includes worlds where **B** fires, we find that F does security-depend on C. In parallel with what we found in section 3.2, then, the difference between the case where **B** fires (Figure 19) and the case where **B** does not fire (Figure 20) is the following: in the case where **B** fires, any possibility horizon that treats C as a candidate cause yields the result that F security-depends on C. By contrast, in the case where **B** does not fire, the verdict on whether F security-depends on C is much more sensitive to our choice of possibility horizon: within the larger possibility horizon, which includes worlds where **B** fires, F still security-

These results allow my account of causation to accommodate and explain our intuitive verdicts on Figure 19 and 20: in Figure 19, *C* is a cause of *F*, since *C* is process-connected to *F* and *F* security-depends on *C* within the relevant possibility horizon. In Figure 20, the intrinsicness of process-connection ensures that *C* is still process-connected to *F*. However, as we have just seen, *F* does not security-depend on *C* within the relevant possibility horizon $\mathscr{H}$ – and for this reason, *C* is not a cause of *F* within $\mathscr{H}$, since the necessary condition of security-dependence fails to be satisfied.

4. Conclusion

In this chapter, we have seen how the condition of security-dependence allows my account to handle three groups of puzzling cases: firstly, it captures the sense in which a cause makes a difference to its effect even in cases of redundant causation. Secondly, it accommodates our intuitive judgements in the counterexamples to the transitivity of causation, and provides a principled way of drawing the line between transitivity-failure cases and early preemption cases. And thirdly, it accommodates and explains our intuitive judgements in the counterexamples to the intrinsicness of causation. In the following chapter, we shall see further examples of how the condition of security-dependence allows us to handle otherwise puzzling cases.

depends on *C*; but within the smaller and more natural choice of possibility horizon, where **B**'s failure to fire is treated as a background condition, we find that *F* does not security-depend on *C* – instead, *C* is simply irrelevant to *F*.

10

More applications of security-dependence

In this chapter, I show how the condition of security-dependence allows us to handle three further kinds of cases. First, I show how it resolves the so-called problem of profligate omissions (section 1). Second, I show how it allows my account to deliver intuitively correct verdicts on structurally isomorphic but causally different cases (section 2). And finally, I show how the condition of security-dependence allows my account to handle contrastive causal claims (section 3).

1. The problem of profligate omissions

In this section, I show how the condition of security-dependence allows my account to handle the following puzzling case (previously discussed in Chapter 2 section 1.3.1):

> *The flowers*: Suzy goes on holiday and Billy promises to water her flowers while she is away. However, Billy does not water the flowers and the flowers die.

Intuitively, Billy's failure to water the flowers is a cause of their death. However, it is also the case that the queen does not water the flowers. And it seems that the relation between the queen's failure to water the flowers and their death is just the same as the relation between Billy's failure to water the flowers and their death. But intuitively, we do not count the queen's failure to water the flowers as a cause of their death. In this section, I show how my proposed account of causation can accommodate this verdict.

The structure of the case is illustrated below:

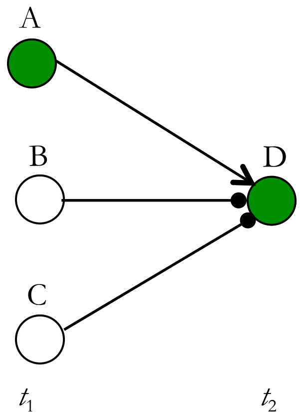


Figure 1

A represents the dry soil in the pot, $\neg B$ represents Billy's failure to water the flowers, $\neg C$ represents the queen's failure to water the flowers, and D represents the death of the flowers. Structurally, the queen's failure to water the flowers is related to their death in exactly the same way as Billy's failure to water them. So why do we count one as a cause, but not the other?

The condition of process-connection cannot on its own capture the difference between Billy's failure to water the flowers and the queen's failure to water the flowers: we find that $\neg B$ (Billy's failure to water the flowers) is process-connected to D via the following process:

t_1:	$\neg B$	(together with A and $\neg C$)
t_2:	D	

This is illustrated below:

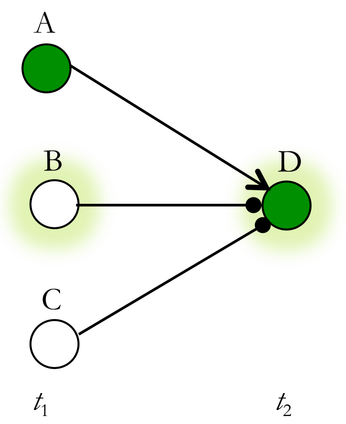


Figure 1*

And we similarly find that $\neg C$ is process-connected to D, via the following process:

t_1: $\neg C$ (together with A and $\neg B$)

t_2: D

This is illustrated below:

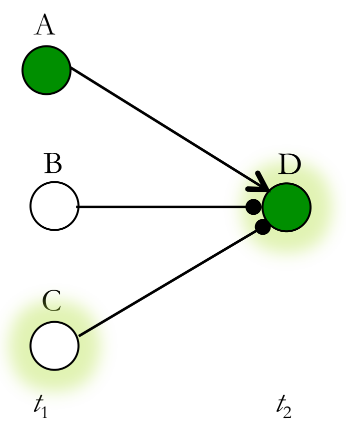


Figure 1**

Thus, the condition of process-connection gives a parallel treatment to Billy's failure to water the flowers and the queen's failure to water the flowers.

To capture the difference, we need the condition of security-dependence: in ordinary contexts, it is natural to treat Billy's failure to water the flowers as a deviant event, but treat the queen's failure to water the flowers as default. This yields a possibility horizon containing just the following four possible worlds:

	Complete state at time t_1:	*Events at t_1:*	t_2:
@:	**A** fires, **B** does not fire, **C** does not fire	A, $\neg B$, $\neg C$	D
w_1:	**A** does not fire, **B** does not fire, **C** does not fire	A, $\neg B$, $\neg C$	D
w_2:	**A** fires, **B** fires, **C** does not fire	A, $\neg B$, $\neg C$	D
w_3:	**A** does not fire, **B** fires, **C** does not fire	A, $\neg B$, $\neg C$	D

Call this possibility horizon $\mathscr{H}$. Within this possibility horizon, $\neg B$, i.e. Billy's failure to water the flowers, is treated as a candidate cause: there are some worlds in $\mathscr{H}$ where $\neg B$ does not occur. By contrast, $\neg C$, i.e. the queen's failure to water the flowers, is treated as a background condition: it is held fixed throughout our possibility horizon. This difference accounts for the fact that $\neg B$ counts as a cause of D within $\mathscr{H}$, while $\neg C$ does not count as a cause of D within $\mathscr{H}$.

First, we find that D depends counterfactually on $\neg B$ within $\mathscr{H}$: from the standpoint of @, the closest-at-t_1 world in $\mathscr{H}$ where $\neg B$ does not occur is w_2. D does not occur in w_2. Thus, D depends counterfactually on $\neg B$ within $\mathscr{H}$. And it follows immediately from this that D security-depends on $\neg B$ within $\mathscr{H}$. Thus, we get the intuitively correct verdict that $\neg B$ is indeed a cause of D within the contextually relevant possibility horizon.

By contrast, we find that D does *not* depend counterfactually on $\neg C$ within $\mathscr{H}$, for the simple reason that there is no world within $\mathscr{H}$ where $\neg C$ does not occur. This already shows that the case presents no counterexample to the principle that counterfactual dependence within a possibility horizon is sufficient for causation within that same possibility horizon.

To reject $\neg C$ as a cause of D, however, we need the stronger result that D does not security-depend on $\neg C$ within $\mathscr{H}$. And we get this result too, for the very same reason – namely, because there is no world within $\mathscr{H}$ where $\neg C$

does not occur. Thus, we get the intuitively correct verdict that $\neg C$ is *not* a cause of D within the contextually relevant possibility horizon, because this possibility horizon simply does not treat it as a relevant possibility that $\neg C$ could fail to occur.

This treatment of the case fits nicely with the nuances of our intuitive judgement that the queen's failure to water the flowers is not a cause of their death: intuitively, we reject the queen's failure to water the flowers as a cause because we do not take seriously the possibility that the queen could have watered the flowers. As Schaffer suggests:

> 'Perhaps the reason it sounds wrong to say that the queen's not watering [the] flowers causes them to wilt is that we never supposed that the queen would deign to water [the] flowers. […] We resist taking such an unrealistic supposition as a contrast.'[1]

Schaffer's suggestion fits precisely with the suggestion I make here: we reject the queen's not watering the flowers as a cause because there is no world in our possibility horizon in which the queen *does* water the flowers. And thus, the queen's failure to water the flowers is treated simply as a background condition, rather than a candidate cause.

The problem discussed here is often understood as a problem about omissions, and is sometimes referred to as the *problem of profligate omissions*.[2] However, the distinction between omissions and so-called positive events does not in fact play any role in generating the problem: essentially the same problem may arise in cases that involve only positive events. Suppose, for example, that there is a lightning strike in a forest and a forest fire starts immediately thereafter. In that case, we would normally count the lightning

[1] Schaffer (2005), p. 302.

[2] See e.g. Bernstein (2014), p. 429. For discussion, see also Beebee (2004); McGrath (2005); Sartorio (2010), pp. 262-3.

strike, but not the presence of oxygen, as a cause of the fire (see Chapter 5 section 4). Thus, a better name would be the *problem of background conditions*.[3]

Correspondingly, the distinction between omissions and positive events plays no role in my proposed solution. Rather, my proposed solution is based on the way in which a possibility horizon encodes the distinction between candidate causes and background conditions: Billy's failure to water the flowers is a candidate cause; the queen's failure to water the flowers is a mere background condition. And similarly in the case of the forest fire: the lightning strike is a candidate cause; the presence of oxygen is a mere background condition.

2. Structurally isomorphic but causally different cases

An important advantage of the condition of security-dependence is that it allows us to accommodate causally different but structurally isomorphic cases. To illustrate this, I will now consider an intriguing case originally presented by McDermott. We may set out this case as follows:

> *Wall and window*: Suzy throws a ball towards a window. Before the ball reaches the window, Billy leaps up and catches the ball. If Billy had not caught the ball, it still would not have hit the window – for between Billy and the window, there is a sturdy brick wall.[4]

The most natural verdict on this case is that Billy's catch is not a cause of the window remaining intact, since the window was not in any danger of being broken either way. As Collins writes:

[3] Cf. Menzies (2004), pp. 142-45.

[4] Cf. McDermott (1995), p. 525. For discussion, see also Collins (2004); Hall (2007*b*), pp. 60-61; Halpern and Pearl (2005), pp. 876-77; Lewis (2004*a*), pp. 102-3; Maudlin (2004), pp. 435-38; and Woodward (2003), pp. 86-91.

> 'I am very reluctant to say in this case that [Billy's] catch prevented the ball from breaking the window. Given that the wall was there, the window was never in any danger of being broken.'[5]

Against this, McDermott makes the following argument: if Billy had not caught the ball and the wall had not been there, the window would have broken. So between them, Billy and the wall prevented the window from breaking. Clearly, however, the wall contributed nothing. So the credit should go to Billy. McDermott reports that on hearing this argument, most people retract their initial judgement and accept Billy's catch as a cause of the window remaining intact.[6]

These conflicting verdicts already make the case quite interesting. It becomes even more interesting once we note that our initial intuitive verdict is reversed when a second catcher replaces the wall, as in the following case:

> *Catcher and window*: Suzy throws a ball towards a window. Before the ball reaches the window, Billy leaps up and catches the ball. If Billy had not caught the ball, it still would not have hit the window – for between Billy and the window stands Sally, who would have caught the ball if Billy had failed to do so.[7]

In this case, it seems intuitively clear that Billy's catch *does* cause the window to remain intact.

The two cases – *Wall and window* and *Catcher and window* – are structurally isomorphic. Indeed, the following neuron diagram, which we have already considered in Chapter 7 section 3.7 (Figure 12), represents the structure of both cases:

[5] Collins (2004), p. 108.
[6] McDermott (1995), p. 525.
[7] See e.g. Collins (2004), p. 107.

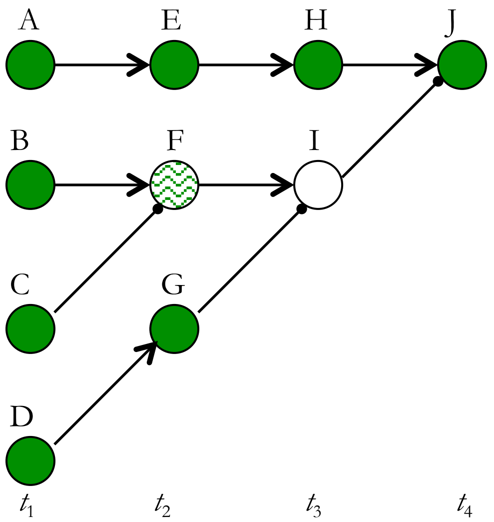


Figure 2

A here represents the presence of the intact window, *B* represents Suzy's throw, *C* represents Billy being ready to catch the ball, and *D* represents the presence of the wall (or Sally's being ready to catch the ball). *F-waves* represents Billy's stopping the ball, ¬*I* represents the absence of a ball moving towards the window, and *J* represents the window's being intact. Given that *Wall and window* and *Catcher and window* are structurally isomorphic, how should we explain the causal difference between them?

The condition of process-connection cannot on its own accommodate the causal difference between the two cases: in either case, we find (as we have seen in Chapter 7 section 3.7) that *C* is process-connected to *J* via the process:[8]

t_1:	*C*	(together with *B*)
t_2:	*F-waves*	
t_3:	¬*I*	(together with *H*)
t_4:	*J*	

This process is illustrated below:

[8] As we have also seen, we find – as we should – that *D* is not process-connected to *J*. This fits the intuitive verdict that *D* is merely a preempted backup.

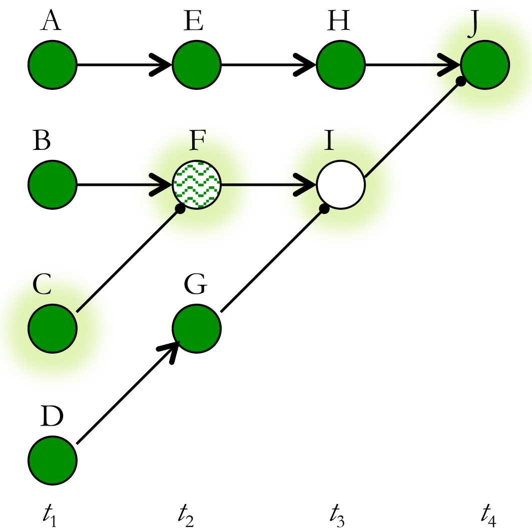


Figure 2*

To explain and accommodate our intuitive verdicts on these cases, we need the condition of security-dependence. As we shall see, this condition can – because it brings in the contextually relevant possibility horizon as a third relatum – explain and accommodate both of our conflicting verdicts on the original case of *Wall and window* as well as our verdict on *Catcher and window*.

In *Wall and window*, it is – at least in normal contexts – reasonable to categorise A (the presence of the intact window) and D (the presence of the sturdy brick wall) as default events: when we extrapolate from what was the case just before t_1, the result of our extrapolation is that the wall and the window are both there. Thus, our possibility horizon should include just the following four worlds, characterised by their complete state at time t_1:

	Complete state at time t_1:	*Events at t_1:*	t_4:
@:	**A** fires, **B** fires, **C** fires, **D** fires	A, B, C, D	J
w_1:	**A** fires, **B** does not fire, **C** fires, **D** fires	A, B, C, D	J
w_2:	**A** fires, **B** fires, **C** does not fire, **D** fires	A, B, C, D	J
w_3:	**A** fires, **B** does not fire, **C** does not fire, **D** fires	A, B, C, D	J

Call this possibility horizon $\mathscr{H}_1$. Note that A (the presence of the intact window) and D (the presence of the wall) are here treated as background conditions, and are held fixed throughout $\mathscr{H}_1$. Note further that it follows from the neuron laws that J (the window's remaining intact) occurs in every world in $\mathscr{H}_1$ – our possibility horizon does not include *any* world in which J fails to occur.

This means that J is maximally secure in every world in $\mathscr{H}_1$. Thus, J (the window's remaining intact) does not security-depend on *anything* within $\mathscr{H}_1$. In particular, J does not security-depend on C (Billy's being ready to catch the ball): J would have been *just as secure* at t_1 if C had not occurred. Thus, we find that C (Billy's being ready to catch the ball) is *not* a cause of J (the window's remaining intact): while C is process-connected to J, we find that J does not security-depend on C within $\mathscr{H}_1$ – and so, C is not a cause of J within $\mathscr{H}_1$, since it fails to satisfy the necessary condition of security-dependence.

Furthermore, this correctly captures the nuances of our intuitive judgement: as we have seen, J (the window's remaining intact) would have been *just as secure* at t_1 if C (Billy's being ready to catch the ball) had not occurred. Correspondingly, we intuitively judge that C (Billy's being ready to catch the ball) is simply *irrelevant* to J (the window's remaining intact).

Our possibility horizon $\mathscr{H}_1$ fits with a way of viewing the case – implicit in Collins' comment that '[g]iven that the wall was there, the window was never in any danger of being broken'[9] – where the window's remaining intact is not in need of causal explanation: J does not security-depend on *anything* within $\mathscr{H}_1$. Since security-dependence is a necessary condition for causation, this means that J does not have *any* causes within $\mathscr{H}_1$. Rather, the breaking of the window is simply treated as *impossible* within $\mathscr{H}_1$, in the sense that $\mathscr{H}_1$ does not contain any world in which the window breaks. This means that the window's remaining intact is not in need of causes: within $\mathscr{H}_1$, the correct response to

[9] Collins (2004), p. 108.

the question of what caused the window to remain intact is a rejection of the question itself – for example, by pointing out that the window was protected by a sturdy brick wall, and so, the window was not in danger in the first place.

To the extent that it is successful, McDermott's argument for saying that Billy's catch caused the window to remain intact shifts the context by forcing us to find a cause of the window's remaining intact. To find such a cause, we need to widen our possibility horizon by categorizing the presence of the wall as a deviant event. For example, extrapolation from the surroundings yields the result that there is just thin air where the wall in fact is. If we accept this as an alternative to the presence of the wall, we now need to include the following eight possible worlds in our possibility horizon:

	Complete state at time t_1:	*Events at t_1:*	t_4:
@:	**A** fires, **B** fires, **C** fires, **D** fires	A, B, C, D	J
w_1:	**A** fires, **B** does not fire, **C** fires, **D** fires	A, B, C, D	J
w_2:	**A** fires, **B** fires, **C** does not fire, **D** fires	A, B, C, D	J
w_3:	**A** fires, **B** does not fire, **C** does not fire, **D** fires	A, B, C, D	J
w_4:	**A** fires, **B** fires, **C** fires, **D** does not fire	A, B, C, D	J
w_5:	**A** fires, **B** does not fire, **C** fires, **D** does not fire	A, B, C, D	J
w_6:	**A** fires, **B** fires, **C** does not fire, **D** does not fire	A, B, C, D	J
w_7:	**A** fires, **B** does not fire, **C** does not fire, **D** does not fire	A, B, C, D	J

Call this possibility horizon $\mathscr{H}_2$. Note that $\mathscr{H}_2$ contains all the same worlds as $\mathscr{H}_1$, but in addition it contains worlds in which D does not occur (i.e. worlds where the wall is replaced with thin air). Within this more inclusive possibility horizon, we find that there *is* a world in which J (the window's remaining intact) does not occur – namely, world w_6.

We may represent the comparative distances-at-t_1 between the worlds in $\mathscr{H}_2$ as follows, where the world in which J does not occur is coloured grey:

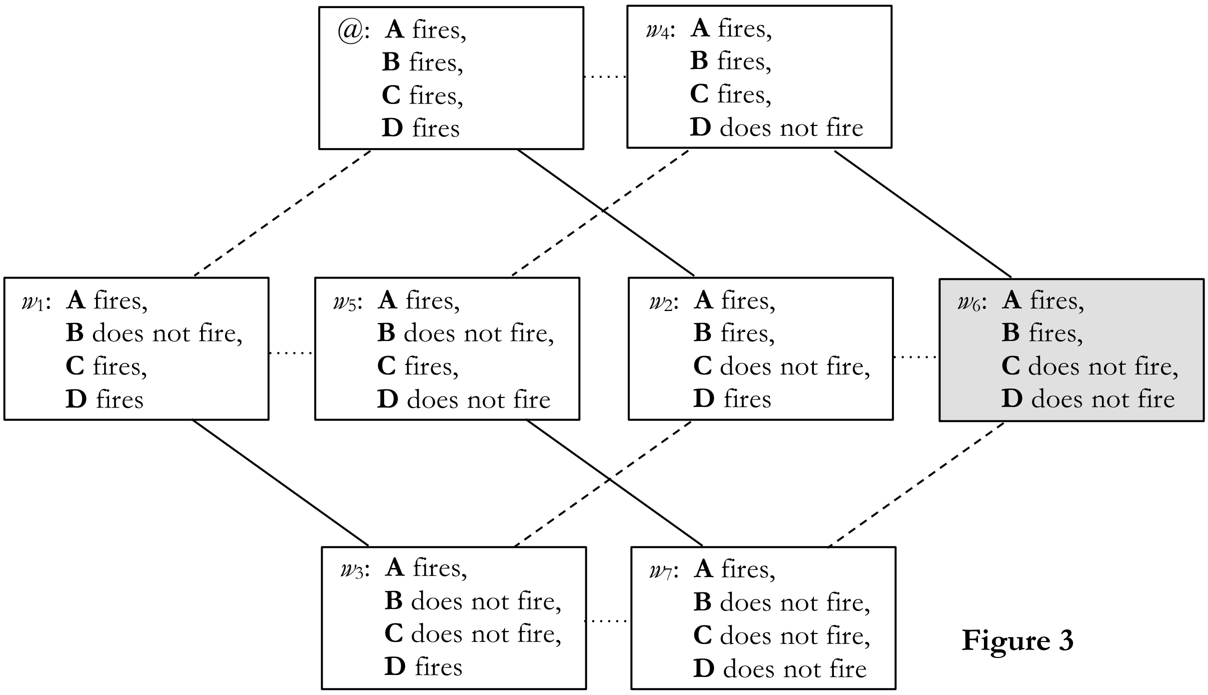


Figure 3

We now find that *J* (the window's remaining intact) security-depends on *C* (Billy's being ready to catch the ball) within $\mathscr{H}_2$. To see this, let us go through our usual three steps:

Step 1: our first step is to evaluate *J*'s security at t_1 in @. From the standpoint of @, there is one minimal dependence set for *J* at t_1 – namely $\{C, D\}$. The closest-at-t_1 world where both *C* and *D* fail to occur is w_6. Thus, *J*'s security at t_1 in @ is given by the distance-at-t_1 from @ to w_6.

Step 2: our second step is to evaluate *J*'s security at t_1 in the closest-at-t_1 world(s) where *C* does not occur. From the standpoint of @, the closest-at-t_1 world where *C* does not occur is w_2. From the standpoint of w_2, *J* has one minimal dependence set, namely $\{D\}$, and the closest-at-t_1 world where *D* does not occur is w_6. Thus, *J*'s security at t_1 in w_2 is given by the distance-at-t_1 from w_2 to w_6.

Step 3: our final step is to compare *J*'s security at t_1 in @ with *J*'s security at t_1 in w_2. As shown in the figure, the distance-at-t_1 from w_2 to w_6 is strictly shorter than the distance-at-t_1 from @ to w_6. Thus, we find that *J* would have been *less secure* at t_1 if *C* had not occurred.

Within $\mathscr{H}_2$, then, we find that *J* (the window's remaining intact) does security-depend on *C* (Billy's being ready to catch the ball). Since *C* is also process-connected to *J*, as we have seen above, my account yields the verdict that *C* is a cause of *J* within $\mathscr{H}_2$ – thereby accommodating the verdict elicited by McDermott's argument. Thus, we may think of McDermott's argument as bringing about a shift in the contextually relevant possibility horizon, such that it becomes relevant to consider worlds where the wall is absent. This fits nicely with Hall's diagnosis of the case:

> 'what McDermott's argument does, for those inclined to treat it charitably, is to introduce a context where the contrast between a wall-present situation and a wall-absent situation becomes salient.'[10]

As Hall also points out, however, it is most natural to simply treat the wall's presence as default – and this explains 'why the McDermott argument has a whiff of sophistry about it'.[11]

We may now compare this with the structurally isomorphic case where the wall is replaced by Sally, who is ready to catch the ball in case Billy should fail. Whereas it is somewhat strained to treat the presence of the wall as deviant, it is perfectly natural to categorise Sally's readiness to catch the ball as deviant. As Hall writes:

> 'In the actual situation, Sally is alert, poised to intercept if Billy doesn't. We might focus on the contrast between that psychological state, and a state she

[10] Hall (2007*b*), p. 61.

[11] Hall (2007*b*), pp. 60-61.

easily could have been in, in which she not only is idle but is disposed to remain so, perceived threats to the window notwithstanding.'[12]

In the case where Sally replaces the wall, the contextually relevant possibility horizon is therefore the more inclusive $\mathscr{H}_2$, in which *D* (Sally's being ready to catch the ball) is categorised as a deviant event, with $\neg D$ (Sally's being disposed to remain idle) as its corresponding alternative. Within this more inclusive possibility horizon, we find – as we have just seen – that *J* (the window's remaining intact) security-depends on *C* (Billy's readiness to catch the ball). In the case where Sally replaces the wall, then, we find that *C* is a cause of *J* within the contextually relevant possibility horizon $\mathscr{H}_2$, since *C* satisfies both of the necessary and jointly sufficient conditions of process-connection and security-dependence.[13]

3. Contrastive causal claims

The surface form of our causal claims is usually binary. In some cases, however, we make contrastive causal claims – adding a contrast to the cause, to the effect, or both. Such contrastive causal claims take the form '*c* rather than *c** causes *e*', '*c* causes *e* rather than *e**', or '*c* rather than *c** causes *e* rather than *e**'. In this section, I show how the condition of security-dependence, where security-dependence is understood as a ternary relation between two instantaneous events and a possibility horizon, allows a straightforward interpretation of such contrastive causal claims.

Let us begin by considering the following case, where it seems that we can only capture the relevant causal facts by means of contrastive claims:

Switch - variation: Suzy is standing by a switch in the railroad tracks. She sees a train approaching in the distance, and flips the switch so that the train travels

[12] Hall (2007*b*), p. 61.

[13] For my treatment of two other structurally isomorphic but causally different cases, see Appendix B case 31 and 32.

down the local track. She could also have flipped the switch so that the train had travelled down the express track, or the broken track. If she had chosen the express track, the train would have arrived at its destination more quickly. If she had chosen the broken track, the train would have derailed.[14]

The neuron diagram below illustrates the structure of this case:

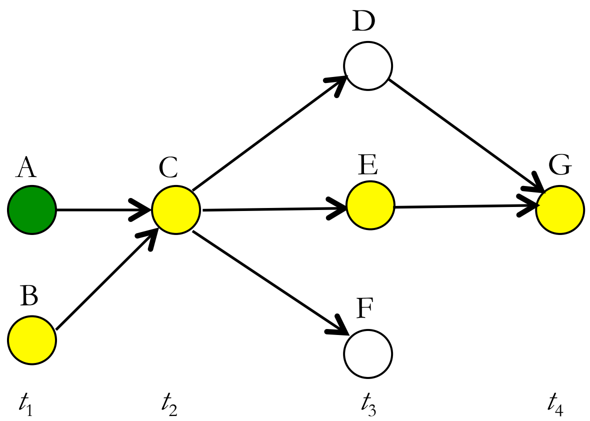


Figure 4

Neuron **B** can here be in three different states: it can fire in green, yellow, or red (it cannot be dormant). Neuron **C** can be in four different states: it can fire in green, yellow, or red, or it can be dormant. **C** fires in green if and only if **A** fires and **B** fires in green; **C** fires in yellow if and only if **A** fires and **B** fires in yellow; and **C** fires in red if and only if **A** fires and **B** fires in red. Finally, **C** is dormant if and only if **A** does not fire. The neuron laws are such that when **C** fires in green, it sends a stimulatory signal to **D** (and no signal to **E** or **F**); when **C** fires in yellow, it sends a stimulatory signal to **E** (and no signal to **D** or **F**); and when **C** fires in red, it sends a stimulatory signal to **F** (and no signal to **D** or **E**). Finally, **G** can fire in green or yellow or remain dormant. **G** fires in green if and only if it receives a stimulatory signal from **D**; **G** fires in yellow if and only if it receives a stimulatory signal from **E**; and otherwise it remains

[14] Schaffer (2012), pp. 38-40. For discussion of this case, see also Steglich-Petersen (2012), pp. 131-32. For discussion of the role of contrasts in causation, see in addition Hitchcock (1996); Maslen (2004); and Northcott (2008).

dormant. *A* here corresponds to the approach of the train, and **B**'s firing in green, yellow, or red corresponds to Suzy's flipping the switch to express, local, or broken, respectively.

The figure below shows what would have happened if **B** had fired in green (i.e. if Suzy had flipped the switch to express):

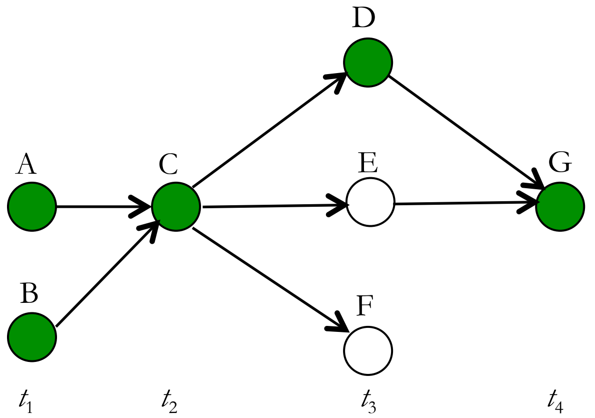


Figure 4*

And finally, the figure below shows what would have happened if **B** had fired in red (i.e. if Suzy had flipped the switch to broken):

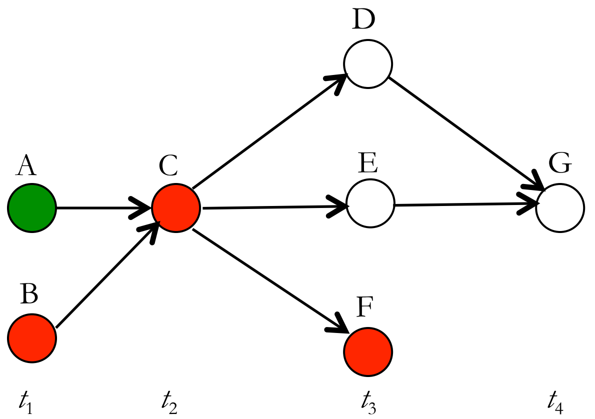


Figure 4**

Let *B-green* be the event that is essentially **B**'s firing in green, let *B-yellow* be the event that is essentially **B**'s firing in yellow, and let *B-red* be the event that is essentially **B**'s firing in red. Furthermore, let *G* be the event that is essentially **G**'s firing (whether in green or yellow), let *G-green* be the event that is essentially **G**'s firing in green, and let *G-yellow* be the event that is essentially **G**'s firing in yellow.

We may now ask: is *B-yellow* (Suzy's flipping the switch to local) a cause of *G* (the train's arrival)? Intuitively, we can only answer this question correctly by specifying a contrast: *B-yellow* (Suzy's flipping the switch to local) rather than *B-green* (Suzy's flipping the switch to express) is *not* a cause of *G* (the train's arrival); but *B-yellow* (Suzy's flipping the switch to local) rather than *B-red* (Suzy's flipping the switch to broken) *is* a cause of *G* (the train's arrival). Thus we find that, in order to answer the question correctly, we need to make a *contrastive* causal claim that explicitly mentions a particular contrast to the cause.

The condition of security-dependence allows us to understand why we need to turn to contrastive claims in this case. To see this, let us begin by considering the following possibility horizon, which includes all nomologically possible worlds:

	Complete state at time t_1:	*Events at* t_1:	t_4:
@:	**A** fires, **B** fires yellow	*A*, *B-green,* *B-yellow*, *B-red*	*G*, *G-green,* *G-yellow*
w_1:	**A** fires, **B** fires green	*A*, *B-green,* *B-yellow*, *B-red*	*G*, *G-green,* *G-yellow*
w_2:	**A** fires, **B** fires red	*A*, *B-green,* *B-yellow*, *B-red*	*G*, *G-green,* *G-yellow*
w_3:	**A** does not fire, **B** fires yellow	*A*, *B-green,* *B-yellow*, *B-red*	*G*, *G-green,* *G-yellow*
w_4:	**A** does not fire, **B** fires green	*A*, *B-green,* *B-yellow*, *B-red*	*G*, *G-green,* *G-yellow*
w_5:	**A** does not fire, **B** fires red	*A*, *B-green,* *B-yellow*, *B-red*	*G*, *G-green,* *G-yellow*

Suppose that the distance-at-t_1 between @ and w_1 is exactly the same as the distance-at-t_1 between @ and w_2, which is exactly the same as the distance-at-t_1 between w_1 and w_2. In that case, we may represent the comparative distance-at-t_1 between these six worlds as follows, where the worlds in which *G* does not occur are coloured grey:

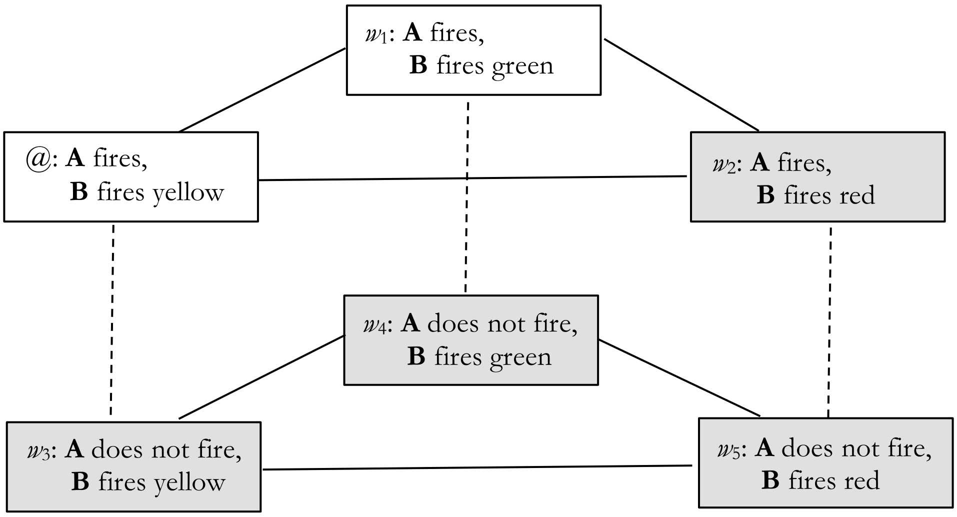


Figure 5

From the standpoint of @, we here find that there are two closest-at-t_1 worlds where *B-yellow* does not occur – namely, w_1 and w_2. Importantly, these two worlds deliver conflicting verdicts on whether *G* would have been *less secure*, *just as secure* or *more secure* at t_1 if *B-yellow* had not occurred. To see this, let us go through our usual three steps:

Step 1: evaluating *G*'s security at t_1 in @. From the standpoint of @, there are two minimal dependence sets for *G* at t_1 – namely, $\{\neg B\text{-}red\}$ and $\{A\}$. The closest-at-t_1 world where $\neg$*B-red* does not occur is w_2, and the closest-at-t_1 world where *A* does not occur is w_3. Thus, *G*'s security at t_1 in @ is given by the distance-at-t_1 between @ and w_2 together with the distance-at-t_1 between @ and w_3.

Step 2: evaluating *G*'s security at t_1 in the closest-at-t_1 world(s) where *B-yellow* does not occur. We already know that these worlds are w_1 and w_2.

From the standpoint of w_1, there are again two minimal dependence sets for *G* at t_1 – namely, $\{\neg B\text{-}red\}$ and $\{A\}$. The closest-at-t_1 world where $\neg$*B-red* does not occur is w_2, and the closest-at-t_1 world where *A* does not occur is w_4. Thus, *G*'s security at t_1 in w_1 is given by the distance-at-t_1 from w_1 to w_2 together with the distance-at-t_1 from w_1 to w_4.

From the standpoint of w_2, there is one minimal dependence set for G at t_1 – namely, the empty set Ø. Thus, G has minimal security at t_1 in w_2.

Step 3: comparing G's security at t_1 in @ with G's security at t_1 in w_1 and w_2. We here find that G is *just as secure* at t_1 in w_1 as it is in @, since the distance-at-t_1 between @ and w_2 is exactly the same as the distance-at-t_1 between w_1 and w_2, and the distance-at-t_1 between @ and w_3 is exactly the same as the distance-at-t_1 between w_1 and w_4. By contrast, we find that G is *less secure* at t_1 in w_2 than it is in @, since G has minimal security at t_1 in w_2.

Thus, our possibility horizon does not yield a uniform verdict on the question of whether G would have been *less secure*, *just as secure* or *more secure* at t_1 if *B-yellow* had not occurred. To get a uniform verdict, we need to restrict the possibility horizon under consideration. This is precisely what we can do by making contrastive claims:[15] as we have already seen (Chapter 5 section 3), the explicit specification of contrasts imposes certain requirements on the possibility horizon under consideration. To recapitulate: the contrastive causal claim '*c* rather than *c** is a cause of *e*' imposes the requirement that our possibility horizon must contain at least one world in which *c* occurs, at least one world in which *c** occurs, and no world in which neither *c* nor *c** occurs. The contrastive claim '*c* causes *e* rather than *e**' imposes the requirement that our possibility horizon must contain at least one world in which *e* occurs, at least one world in which *e** occurs, and no world in which neither *e* nor *e** occurs. And a quaternary contrastive claim, '*c* rather than *c** is a cause of *e* rather than *e**', imposes both of these requirements at once.

My suggestion is that we explicitly mention contrasts in just those cases where we need to artificially narrow down the possibility horizon under consideration in order to get uniform results.[16] The role of the contrasts is to specify the relevant possibility horizon. In this way, contrastive causal claims express a ternary causal relation – *c* causes *e* within possibility horizon $\mathscr{H}$ – in

---

[15] For a very similar suggestion, see Northcott (2008), pp. 119-20.

[16] Cf. Northcott (2008), pp. 119-20.

exactly the same way as ordinary, binary causal claims. And my account of causation thus straightforwardly applies to contrastive causal claims as well.

In the following, I will consider a number of cases that demonstrate how my account of causation yields intuitively correct verdicts when applied to contrastive causal claims. We may begin by noting that process-connection is a binary relation: *c* is process-connected to *e*. This means that our choice of possibility horizon – and, by extension, our choice of contrasts – has no bearing on whether the condition of process-connection is satisfied. For example, we find that *B-yellow* is process-connected to *G* via the following process – regardless of whether *B-yellow* is contrasted with *B-green* or with *B-red*:

t_1:	*B-yellow*	(together with *A*)
t_2:	*C-yellow*	
t_3:	*E*	
t_4:	*G*	

To explain why our choice of contrasts matters for the truth-value of our causal claims, we need the condition of security-dependence: security-dependence is a ternary relation between two instantaneous events and a possibility horizon. And thus, it may easily happen that *e* security-depends on *c* within a possibility horizon $\mathscr{H}_1$, selected by one choice of contrasts, while *e* does not security-depend on *c* within a different possibility horizon $\mathscr{H}_2$, selected by a different choice of contrasts. This is precisely what happens in the case of *B-yellow*:

Consider first the claim that '*B-yellow* rather than *B-green* is a cause of *G*'. To evaluate this claim, we need to restrict our possibility horizon, so that it contains at least one world where *B-yellow* occurs, at least one world where *B-green* occurs, and no world where neither *B-yellow* nor *B-green* occurs. The largest possibility horizon that satisfies these requirements contains the following four worlds:

	Complete state at time t_1:	*Events at t_1*:	t_4:
@:	**A** fires, **B** fires yellow	*A*, *B-green, B-yellow*, *B-red*	*G*, *G-green, G-yellow*
w_1:	**A** fires, **B** fires green	*A*, *B-green, B-yellow*, *B-red*	*G*, *G-green, G-yellow*
w_3:	**A** does not fire, **B** fires yellow	*A*, *B-green, B-yellow*, *B-red*	*G*, *G-green, G-yellow*
w_4:	**A** does not fire, **B** fires green	*A*, *B-green, B-yellow*, *B-red*	*G*, *G-green, G-yellow*

Within this restricted possibility horizon, we now find that *G* does not security-depend on *B-yellow*. Rather, *B-yellow* is simply *irrelevant* to *G*: *G* would have been *just as secure* at t_1 if *B-yellow* had not occurred.[17]

By contrast, the causal claim '*B-yellow* rather than *B-red* is a cause of *G*' imposes the following requirement on our possibility horizon: it must contain at least one world in which *B-yellow* occurs, at least one world in which *B-red* occurs, and no world in which neither *B-yellow* nor *B-red* occurs. The largest possibility horizon that satisfies these requirements is the following:

	Complete state at time t_1:	*Events at t_1*:	t_4:
@:	**A** fires, **B** fires yellow	*A*, *B-green, B-yellow*, *B-red*	*G*, *G-green, G-yellow*
w_2:	**A** fires, **B** fires red	*A*, *B-green, B-yellow*, *B-red*	*G*, *G-green, G-yellow*
w_3:	**A** does not fire, **B** fires yellow	*A*, *B-green, B-yellow*, *B-red*	*G*, *G-green, G-yellow*
w_5:	**A** does not fire, **B** fires red	*A*, *B-green, B-yellow*, *B-red*	*G*, *G-green, G-yellow*

Within this possibility horizon, we find that *G* does security-depend on *B-yellow*. Indeed, *G* depends counterfactually on *B-yellow*: if *B-yellow* had not occurred, *B-red* would have occurred instead – and in that case, *G* would not have occurred.

In this way, the condition of security-dependence allows us to capture why our choice of contrasts matters: regardless of our choice of contrasts, we find that *B-yellow* is process-connected to *G*. However, our choice of contrasts plays a role in determining the relevant possibility horizon: when *B-yellow* is

[17] Note that we get the same result when we consider an even more restricted possibility horizon that contains only @ and w_1.

contrasted with *B-green*, this selects a possibility horizon in which *G* does not security-depend on *B-yellow*, and for this reason the causal claim '*B-yellow* rather than *B-green* causes *G*' comes out false. But when *B-yellow* is contrasted with *B-red*, this selects a possibility horizon in which *G* does security-depend on *B-yellow*, and thus the causal claim '*B-yellow* rather than *B-red* causes *G*' comes out true.

We may treat causal claims that explicitly mention contrasts to the effect in a similar way: these too may be understood as specifications of the relevant possibility horizon. As an example, consider the contrastive claim: '*A* caused *G-yellow* rather than *G-green*'. This claim imposes the requirement that our possibility horizon must contain at least one world in which *G-yellow* occurs, at least one world in which *G-green* occurs, and no world in which neither *G-yellow* nor *G-green* occurs. This yields the following possibility horizon:

	Complete state at time t_1:	*Events at* t_1:	t_4:
@:	**A** fires, **B** fires yellow	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_1:	**A** fires, **B** fires green	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*

Within this possibility horizon, *A* is simply treated as a background condition. Thus, we find that *G-yellow* does not security-depend on *A* for the simple reason that the relevant possibility horizon contains no worlds where *A* does not occur. And so, we find – as we should – that the contrastive claim '*A* caused *G-yellow* rather than *G-green*' is false.

By contrast, we find that the causal claim '*B-yellow* caused *G-yellow* rather than *B-green*' is true: to evaluate this claim, we need to consider just the same possibility horizon as above. Here, however, we find that *G-yellow* does security-depend on *B-yellow*. Indeed, *G-yellow* depends counterfactually on *B-yellow*: if *B-yellow* had not occurred, *G-yellow* would not have occurred. Once again, this is the intuitively correct result.

Finally, we may now consider quaternary contrastive claims that explicitly mention contrasts on both the cause-side and the effect-side. My suggestion is that these claims should be treated in exactly the same way as the claims we have considered above: to evaluate the quaternary contrastive claim 'c rather than c^* causes e rather than e^*', we need to identify a possibility horizon that contains at least one world where c occurs, at least one world where c^* occurs, at least one world where e occurs, and at least one world where e^* occurs, and that contains no world in which neither c nor c^* occurs, and no world in which neither e nor e^* occurs.

It is easy to see that for some choices of c, c^*, e, and e^*, it may simply be impossible to find a possibility horizon that satisfies all these requirements. In such cases, we intuitively register that the chosen contrasts are in some way *mismatched*, and the contrastive claim misfires. An example of this is the quaternary contrastive claim '*B-yellow* rather than *B-red* caused *G-yellow* rather than *G-green*'. For ease of reference, I repeat the complete list of nomologically possible worlds here:

	Complete state at time t_1:	*Events at t_1*:	*t_4*:
@:	**A** fires, **B** fires yellow	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_1:	**A** fires, **B** fires green	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_2:	**A** fires, **B** fires red	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_3:	**A** does not fire, **B** fires yellow	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_4:	**A** does not fire, **B** fires green	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_5:	**A** does not fire, **B** fires red	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*

The cause-contrast requires us to choose a possibility horizon that excludes w_1 and w_4, since neither *B-yellow* nor *B-red* occurs in these worlds. However, the effect-contrast requires us to choose a possibility horizon that includes world w_1, since this is the only world in which *G-green* occurs. Thus, it is simply impossible to find a possibility horizon that accommodates both the cause-contrast and the effect-contrast. This fits nicely with the intuition that there is something off about the causal claim that '*B-yellow* rather than *B-red* caused *G-yellow* rather than *G-green*': intuitively, this claim simply cannot be evaluated, since it suffers from a mismatch of contrasts.

When the contrasts are appropriately matched, however, it is straightforward to evaluate quaternary contrastive claims. Consider for example the claim '*B-yellow* rather than *B-red* caused *G-yellow* rather than $\neg G$'. The cause-contrast here requires us to include at least one of @ and w_3; include at least one of w_2 and w_5; and exclude w_1 and w_4; while the effect-contrast requires us to include @; include at least one of w_2, w_3, w_4, and w_5; and exclude w_1. These demands are fully compatible. The largest possibility horizon that satisfies them contains the following four worlds:

	Complete state at time t_1:	*Events at t_1*:	t_4:
@:	**A** fires, **B** fires yellow	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_2:	**A** fires, **B** fires red	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_3:	**A** does not fire, **B** fires yellow	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*
w_5:	**A** does not fire, **B** fires red	*A*, *B-green*, *B-yellow*, *B-red*	*G*, *G-green*, *G-yellow*

And within this possibility horizon, we find – as we should – that *G-yellow* security-depends on *B-yellow*. Indeed, *G-yellow* depends counterfactually on *B-yellow*: if *B-yellow* had not occurred, *G-yellow* would not have occurred. And thus, the claim comes out true.

As these examples show, the condition of security-dependence – understood as a ternary relation between two events and a possibility horizon – has the resources to straightforwardly accommodate contrastive causal claims.

4. Conclusion

In this chapter, we have seen further examples of how the condition of security-dependence allows us to accommodate our intuitive verdicts in puzzling cases. In particular, we have seen how the condition of security-dependence resolves the so-called problem of profligate omissions, captures our intuitive verdicts in McDermott's puzzling case of redundant prevention, and allows my account to straightforwardly accommodate contrastive causal claims. These applications provide further reason for accepting security-dependence as a necessary condition for causation.

PART V

The biconditional

11

Principles of causal inference

My proposal is that the two conditions of process-connection and security-dependence are individually necessary and jointly sufficient for causation. Thus, my proposed account of causation may be stated as follows:

> *Causation*: *c* is a cause of *e* within a possibility horizon $\mathscr{H}$ iff
> a) *c* is process-connected to *e*, and
> b) *e* security-depends on *c* within $\mathscr{H}$.

So far, we have seen how this account of causation delivers intuitively correct verdicts on a wide range of specific cases. There is, however, an important further component to our use of 'cause': our practices of causal reasoning. In this chapter, we shall consider how my completed account of causation accommodates our core practices of causal reasoning.

In particular, we shall consider the three intuitive principles of causal inference mentioned in Chapter 2, concerning the sufficiency of counterfactual dependence for causation, and the transitivity and intrinsicness of causation. A preliminary statement of these three principles goes as follows (the asterisks indicate that I do not endorse the principles):

> **Sufficiency of counterfactual dependence for causation*:
> if *e* depends counterfactually on *c*, then *c* causes *e*.

> **Transitivity of causation*: if there is a set of events $\{d_1, d_2, \ldots, d_n\}$,
> such that *c* causes d_1, d_1 causes d_2, . . . , and d_n causes *e*, then *c* causes *e*.

> **Intrinsicness of causation*: if *c* causes *e*, and a structure of events $\mathscr{S}$, including all the events that are involved in a process connecting *c* to *e*, is governed by the same laws and exactly matches a structure of events $\mathscr{S}^*$, then the counterpart c^* of *c* in $\mathscr{S}^*$ is a cause of the counterpart e^* of *e* in $\mathscr{S}^*$.

We have already seen that there are counterexamples to each of these three intuitive principles (Chapter 2 section 1.3). My account of causation sides with our intuitive judgements on these counterexamples. Thus, my account cannot sanction the above principles of causal inference in their unrestricted form. As we shall see, however, my account of causation does sanction *restricted* versions of the three principles:

Each of the above three principles applies to one (but not the other) of the two necessary and jointly sufficient conditions for causation: counterfactual dependence within a possibility horizon is sufficient for security-dependence within that same possibility horizon, but is not sufficient for process-connection; and process-connection is transitive and intrinsic to a process, while security-dependence is not. As a result, my account of causation entails restricted versions of the three principles. And although these restricted versions are of course less powerful than their unrestricted cousins, we shall see that they are still powerful enough to legitimise most of our usual practices of causal reasoning.

It is important to note that the three principles play a very different role in my account compared with other accounts of causation: typically, one or more of the three unrestricted principles is brought in, as an independently motivated principle, to *extend the reach* of an account of causation, allowing it to recognise causes that it otherwise would not. In Lewis's original account of causation, for example, the starting point is that counterfactual dependence is sufficient for causation. The principle that causation is transitive is then brought in, as an independently motivated principle, to extend the reach of the account: by relying on the transitivity of causation, Lewis's account is able to accommodate our intuitive judgements in cases of early preemption, even though the effect does not depend counterfactually on its cause in such cases.[1]

By contrast, the three restricted principles are *logical consequences* of my account. This means that the three restricted principles do not extend the reach of my account: any result that can be obtained by applying one or more of the three restricted principles can also be obtained through a direct application of the two conditions of process-connection and security-dependence. The principles simply work as *shortcuts* in the application of the account.

In the following, I present the three restricted principles in detail, and show how they support our ordinary practices of causal reasoning. In section 1, I discuss the sufficiency of counterfactual dependence for causation; in section 2, I discuss the transitivity of causation; and in section 3, I discuss the intrinsicness of causation.

1. Counterfactual dependence

In this section, I show how my account of causation supports a restricted principle of sufficiency of counterfactual dependence for causation (section 1.1). I then show how this principle yields the result that counterfactual dependence is straightforwardly sufficient for causation when we are dealing with simple neuron diagrams, where each neuron can be in one of two states – either firing or dormant (section 1.2).

1.1 A restricted principle of sufficiency of counterfactual dependence

The intuitive principle that counterfactual dependence is sufficient for causation licenses us to make a particular kind of inference. This is brought out in the following statement of the principle (the asterisk indicates that I do not endorse the principle):

---

[1] Lewis (1986*d*).

> * *Sufficiency of counterfactual dependence for causation*:
> if *e* depends counterfactually on *c*, then *c* causes *e*.

We have already seen that there are counterexamples to this principle. For example, the following case (discussed in Chapter 2 section 1.3.1) presents a counterexample:

> *Scarlet*: The pigeon Sophia has been conditioned to peck at scarlet to the exclusion of all other colours. She is presented with a scarlet triangle and pecks at it.[2]

In this case, Sophia's pecking depends counterfactually on the triangle's being red: if the triangle had not been red, she would not have pecked at it. But intuitively, the triangle's being red is not a cause of Sophia's pecking. Rather, the cause is the more fragile event of the triangle's being *scarlet*.

The condition of process-connection allows my account to side with our intuitive verdict on this case: the triangle's being red is *not* process-connected to Sophia's pecking (see Chapter 7 section 4.1). And since process-connection is a necessary condition for causation, it follows that the triangle's being red is not a cause of Sophia's pecking. However, this also means that my proposed account of causation yields the result that counterfactual dependence is not sufficient for causation. On my proposed account, therefore, the unrestricted principle of *Sufficiency of counterfactual dependence for causation* is false.

However, we cannot simply leave it at that: the principle that counterfactual dependence is sufficient for causation plays a central role in our ordinary practices of causal reasoning. And in recognition of this, the principle is often given a central role in philosophical accounts of causation. As Paul and Hall write:

---

[2] This case is closely based on a case presented in Yablo (1992*a*), p. 257. For similar cases, see Yablo (1992*b*), p. 415, and Sartorio (2010), pp. 266-69.

> 'as a sufficient condition on causation, it [counterfactual dependence] has struck many philosophers as exactly right – and therefore as an excellent starting point for a full-blown analysis of causation.'[3]

It would thus be a significant departure from our normal use of 'cause' if we had to reject all inferences relying on the sufficiency of counterfactual dependence for causation.

Fortunately, we do not need to reject these inferences across the board: as I will now show, my account of causation supports a *restricted* principle of sufficiency of counterfactual dependence for causation, which allows us to maintain many of our ordinary practices of causal reasoning.

As I prove in Appendix A section 4, counterfactual dependence within a possibility horizon entails security-dependence within that same possibility horizon:

> *Sufficiency of counterfactual dependence for security-dependence*:
> if e depends counterfactually on c within a possibility horizon $\mathscr{H}$,
> then e security-depends on c within $\mathscr{H}$.

This immediately shows that process-connection together with counterfactual dependence within a possibility horizon $\mathscr{H}$ is sufficient for causation within $\mathscr{H}$.

However, one might well hope for an even closer connection between counterfactual dependence and causation. I believe that there is indeed such a connection, although it will take a bit of work to bring it out. To motivate what I am about to propose, it will be useful to consider the following case:

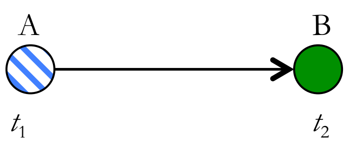


Figure 1

[3] Paul and Hall (2013), p. 16.

Neuron **A** can here fire in four different ways: i) in blue stripes, ii) in uniform blue, iii) in green stripes, and iv) in uniform green. In fact, it fires in blue stripes. Furthermore, the neuron laws are such that **B** fires if and only if **A** fires in either blue stripes or green stripes.

Let *A* be the event that is essentially **A**'s firing (in any of the four ways), let *A-blue* be the event that is essentially **A**'s firing in blue (whether uniform or striped), let *A-stripes* be the event that is essentially **A**'s firing in stripes (whether blue or green), and let *A-blue-stripes* be the event that is essentially **A**'s firing in blue stripes. We then find that there is exactly one process leading up to *B*, namely the following:

t_1:	*A-stripes*
t_2:	*B*

However, if we choose a suitably impoverished possibility horizon, we find that *B* depends counterfactually on each of the four events that occur at time t_1. To see this, consider the following possibility horizon that includes just two worlds:

	Complete state at time t_1:	*Events at t_1:*	t_2:
@:	**A** fires in blue stripes	*A*, *A-blue*, *A-stripes*, *A-blue-stripes*	*B*
w_1:	**A** does not fire	*A*, *A-blue*, *A-stripes*, *A-blue-stripes*	*B*

Call this possibility horizon $\mathscr{H}_1$. We now find that *B* depends counterfactually on *A*, on *A-blue*, on *A-stripes*, and on *A-blue-stripes* within $\mathscr{H}_1$. Thus, *B* depends counterfactually on an event – namely, *A-blue* – that does not even stand in any interesting logical relationship to *A-stripes*, even though *A-stripes* is the only event at t_1 that is process-connected to *B*.

This shows that there is not much of a connection between process-connection and counterfactual dependence within a possibility horizon. And on closer inspection, this should perhaps not come as a surprise: according to my definition, *any* class of nomologically possible worlds that includes the actual world counts as a possibility horizon. It is therefore to be expected that we cannot gain much information about process-connection from the fact that *e* depends counterfactually on *c* within *some* possibility horizon.

To establish a closer connection between counterfactual dependence and process-connection, we need to ensure that the possibility horizon under consideration satisfies more demanding requirements. In particular, the possibility horizon under consideration needs to be *discriminating* at the time when the candidate cause occurs:

> *Discriminating*: a possibility horizon $\mathscr{H}$ is *discriminating* at a time *t* iff for any event *c* occurring in @ at *t*, it is the case that either
>
> a) *c* occurs in every world in $\mathscr{H}$, or
>
> b) *c* and more or less fragile versions of *c* are the only events that occur in @ at *t*, but do not occur in the closest-at-*t* not-*c*-world(s) in $\mathscr{H}$.

This requirement rules out impoverished possibility horizons such as $\mathscr{H}_1$ above. To see that $\mathscr{H}_1$ is not discriminating at time t_1, consider for example the event *A-stripes*, which occurs in @ at t_1. The closest-at-t_1 world in $\mathscr{H}_1$ where *A-stripes* does not occur is w_1. The events that occur in @ at t_1 and do not occur in w_1 are *A*, *A-blue*, *A-stripes*, and *A-blue-stripes*. Obviously, *A-blue* is not a more or less fragile version of *A-stripes*. And so we find that $\mathscr{H}_1$ is not discriminating at time t_1.

By contrast, the following larger possibility horizon *is* discriminating at t_1:

	Complete state at time t_1:	*Events at* t_1:	t_2:
@:	**A** fires in blue stripes	*A*, *A-blue*, *A-stripes*, *A-blue-stripes*	*B*
w_1:	**A** does not fire	*A*, *A-blue*, *A-stripes*, *A-blue-stripes*	*B*
w_2:	**A** fires in uniform blue	*A*, *A-blue*, *A-stripes*, *A-blue-stripes*	*B*
w_3:	**A** fires in green stripes	*A*, *A-blue*, *A-stripes*, *A-blue-stripes*	*B*
w_4:	**A** fires in uniform green	*A*, *A-blue*, *A-stripes*, *A-blue-stripes*	*B*

Call this possibility horizon $\mathscr{H}_2$. We now find that counterfactual dependence within a discriminating possibility horizon, such as $\mathscr{H}_2$, is more revealing: suppose that the distance-at-t_1 between @ and w_2 is strictly shorter than the distance-at-t_1 between @ and w_3. In that case, we find that within $\mathscr{H}_2$, *B* depends counterfactually on three events occurring at time t_1 – namely, *A*, *A-stripes*, and *A-blue-stripes*. Of these, *A-stripes* is process-connected to *B*, and *A* and *A-blue-stripes* stand in an interesting logical relationship to *A-stripes*: *A-stripes* is a less fragile version of *A-blue-stripes*, and *A-stripes* is a more fragile version of *A*. This suggests the following restricted principle of sufficiency of counterfactual dependence for process-connection:

> *Restricted sufficiency of counterfactual dependence for process-connection*:
> if *e* depends counterfactually on *c* within a possibility horizon $\mathscr{H}$,
> such that $\mathscr{H}$ is discriminating at *c*'s time of occurrence,
> then *c* or a more or less fragile version of *c* is process-connected to *e*.

I give a proof of this result in Appendix A section 1, given an auxiliary assumption that is satisfied in all ordinary cases, including all neuron diagrams. Given this, we immediately get the following restricted principle of sufficiency of counterfactual dependence for causation:

> *Restricted sufficiency of counterfactual dependence for causation*:
> if *e* depends counterfactually on *c* within a possibility horizon $\mathscr{H}$,
> such that $\mathscr{H}$ is discriminating at *c*'s time of occurrence,
> then *c* or a more or less fragile version of *c* is a cause of *e* within $\mathscr{H}$.

In the following section, I show how this principle entails that counterfactual dependence is straightforwardly sufficient for causation when we are dealing with simple neuron diagrams.

1.2 Applications

The above principle is particularly easy to apply when we are dealing with simple neuron diagrams, such as the neuron diagram below, which shows a standard case of double prevention:

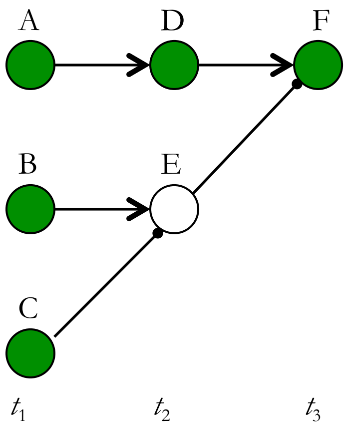


Figure 2

In simple neuron diagrams such as this, each neuron can be in just one of two states – firing or dormant, and we typically restrict the range of our quantifiers to include only maximally temporally fragile events based on the behaviour of a single neuron – i.e., events such as *A*, *B*, *C*, *D*, ¬*E*, etc. This ensures that any way of categorising our neuron events as either default or deviant events, with corresponding alternatives, yields a discriminating possibility horizon. And since our domain of neuron events only includes events at a single level of detail, this means that, in dealing with neuron diagrams such as these, counterfactual dependence within a possibility horizon is straightforwardly sufficient for causation within that same possibility horizon.

Considering the case in Figure 2, for example, we find that *F* depends counterfactually on *C* within any possibility horizon that treats *C* as a candidate cause. The figure below illustrates what happens in the closest-at-t_1 not-*C*-world within any such possibility horizon:

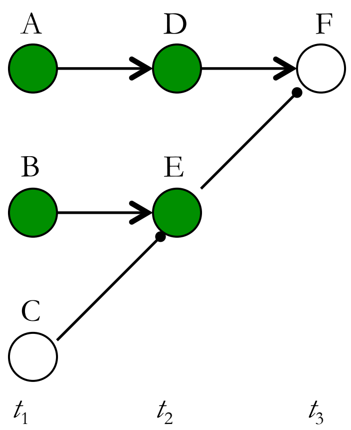


Figure 2*

We thus immediately get the result that *C* is a cause of *F*. This is how we *should* reason in accordance with my proposed account of causation. And it is also how we in fact *do* reason when faced with cases of this kind.

2. Transitivity

The second of the three general principles cited in the introduction is that causation is transitive. I begin by showing how my account of causation entails a restricted principle of transitivity (section 2.1). I then illustrate the applications of this principle (section 2.2).

2.1 A restricted principle of transitivity

We may understand the principle that causation is transitive as licensing a particular kind of inference. To bring this out, we may state the principle as follows (the asterisk indicates that I do not endorse the principle):

> * *Transitivity of causation*: if there is a set of events, $\{d_1, d_2, \ldots, d_n\}$, such that c causes d_1, d_1 causes d_2, . . . , and d_n causes e, then c causes e.[4]

[4] For the sake of simplicity, I do not explicitly mention possibility horizons in this statement.

We have already seen that there are counterexamples to *Transitivity of causation* (Chapter 2 section 1.3.2). We have also seen that my proposed account of causation sides with our intuitive judgements about these counterexamples – and thus goes against *Transitivity of causation* (Chapter 9 section 2). On my proposed account of causation, then, *Transitivity of causation* is false.

However, we cannot simply leave it at that, since appeals to transitivity play a crucial role in our ordinary causal reasoning. As Paul and Hall write:

> 'it often seems perfectly appropriate to justify a claim that *C* causes *E* by tracing a causal chain from the first event to the second'[5]

Indeed, the role of transitivity in our causal reasoning has often been taken as the main argument in favour of holding that causation itself is transitive. For example, Hall has given the following argument:

> 'rejecting transitivity seems intuitively wrong: it *certainly* goes against one of the ways in which we commonly justify causal claims. That is, we often claim that one event is a cause of another precisely *because* it causes an intermediate, which then causes another intermediate, . . . which then causes the effect in question. Are we to believe that any such justification is fundamentally misguided?'[6]

We find a particularly striking illustration of the role that transitivity can play in our causal reasoning when we consider the role of transitivity in accounts of causation. Consider, for example, Lewis's original account of causation.[7] Lewis here takes counterfactual dependence to be sufficient for causation. By bringing in the principle that causation is transitive, he is then able to extend the reach of his analysis, so that it can handle cases of early preemption:

---

[5] Paul and Hall (2013), p. 219.

[6] Hall, in an earlier draft of Hall (2004). The passage is quoted in Maslen (2004), p. 349. See also Schaffer (2005), p. 309, who notes that '[t]he transitive inference feels virtually analytic.'

[7] Lewis (1986*d*).

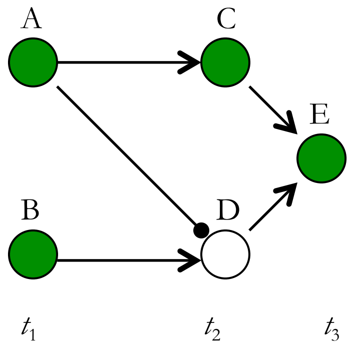


Figure 3

Lewis's theory handles this case as follows: *C* depends counterfactually on *A*, and so *A* is a cause of *C*. Furthermore, *E* depends counterfactually on *C*, and so *C* is a cause of *E*. By transitivity, it then follows – as desired – that *A* is a cause of *E*.

On my proposed account of causation we do not *need* to appeal to transitivity to get this result: we can simply apply the conditions of process-connection and security-dependence directly. However, the above reasoning still seems to provide a persuasive *argument* for holding that *A* is a cause of *E*. If we reject *Transitivity of causation* without providing a substitute, this argument turns out to be entirely unfounded. That will not do. What we need, rather, is a *restricted* principle of transitivity, showing when appeals to transitivity are legitimate and when they are not. As Paul and Hall write:

> 'What's needed is a more developed story, according to which the inference from "*C* causes *D*" and "*D* causes *E*" to "*C* causes *E*" is safe, provided such-and-such conditions obtain'[8]

As I will now show, my account of causation supports just such a restricted principle of transitivity. The key to this restricted principle is to note that the relation of process-connection is transitive, as I prove in Appendix A section 2:

[8] Paul and Hall (2013), p. 219

> *Transitivity of process-connection*: if there is a set of events, $\{d_1, d_2, \ldots, d_n\}$, such that c is process-connected to d_1, d_1 is process-connected to $d_2, \ldots,$ and d_n is process-connected to e, then c is process-connected to e.

From this, it immediately follows that my account of causation supports the following restricted principle of transitivity:

> *Restricted transitivity of causation*: if there is a set of events, $\{d_1, d_2, \ldots, d_n\}$, such that c causes d_1, d_1 causes $d_2, \ldots,$ and d_n causes e, then c causes e within a possibility horizon $\mathscr{H}$, *provided that e security-depends on c within* $\mathscr{H}$.

It is easy to verify that my account of causation entails this principle: from the fact that c causes d_1, it follows that c is process-connected to d_1; from the fact that d_1 causes d_2, it follows that d_1 is process-connected to $d_2, \ldots,$ and from the fact that d_n causes e, it follows that d_n is process-connected to e. From the *Transitivity of process-connection*, it now follows that c is process-connected to e. This shows that c causes e within a possibility horizon $\mathscr{H}$, provided that e security-depends on c within $\mathscr{H}$.

This restricted principle of transitivity licenses us to appeal to transitivity in exactly those cases where such an appeal does not lead us astray – that is, in exactly those cases where the relation of security-dependence behaves transitively. And as we shall now see, this still allows transitivity to play a central role in our causal reasoning.

2.2 Applications

To illustrate how the restricted principle of transitivity works, I begin by considering our standard case of early preemption. I then turn to a much more complicated case that illustrates the power of our restricted transitivity principle. Consider our standard case of early preemption:

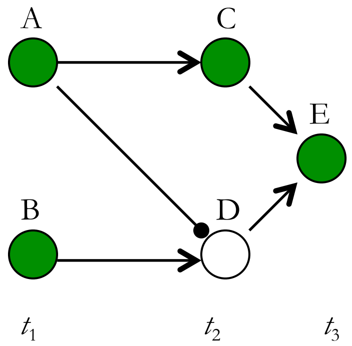


Figure 3

As we have already seen (Chapter 9 section 1.1), *E* security-depends on *A* (and also on *B*) within the contextually relevant possibility horizon. Thus, the restriction on our principle of transitivity is satisfied – and we are now free to appeal to transitivity to show that *A* causes *E*: *C* depends counterfactually on *A*. By our restricted principle of sufficiency of counterfactual dependence for causation, it follows that *A* causes *C*. And *E* depends counterfactually on *C*. So by the same principle, we find that *C* causes *E*. By the restricted transitivity of causation, we now straightforwardly get the desired result that *A* causes *E*. Our restricted transitivity principle thus licenses us to appeal to transitivity in this case, in precisely the way in which we saw Lewis doing so above.

Next, consider the following more complex case:[9]

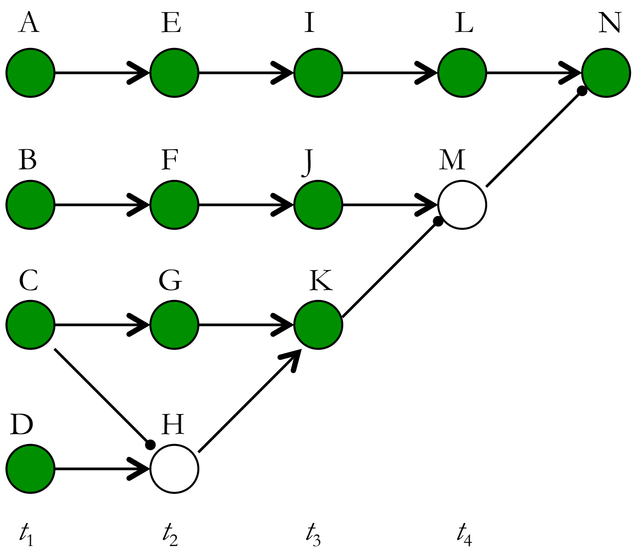


Figure 4

[9] Cf. Paul and Hall (2013), Figure 44.

Is C a cause of N? We begin by noting that N security-depends on C (and also on D) within the possibility horizon $\mathscr{H}$ that treats A, B, C, and D as candidate causes (we get the same result when we consider smaller possibility horizons, provided C and D are treated as candidate causes). Thus, the restriction on our principle is satisfied, and we are free to proceed: by applying the restricted principle of sufficiency of counterfactual dependence for causation, we find that C causes G, G causes K, K causes $\neg M$, and $\neg M$ causes N. By our restricted transitivity principle, it now follows that C causes N within our possibility horizon $\mathscr{H}$.

In this way, our restricted transitivity principle – together with our restricted principle of sufficiency of counterfactual dependence for causation – allows us to arrive at the correct result in a straightforward way, even though we are here faced with a relatively complex case.

3. Intrinsicness

In this section, I turn to the final intuitive principle mentioned in the introduction, namely the principle that causation is intrinsic to a process. I begin by setting out in more detail how we should understand this principle (section 3.1). Next, I show how we may arrive at a restricted intrinsicness principle (section 3.2). And finally, I discuss applications of the restricted principle (section 3.3) and the close connection between idealisation and the restricted intrinsicness principle (section 3.4).

3.1 Understanding the intrinsicness principle

We may give the following preliminary statement of the principle that causation is intrinsic to a process (the asterisk indicates that I do not endorse the principle):

> **Intrinsicness of causation*: if c causes e, and a structure of events $\mathscr{S}$, including all the events that are involved in a process connecting c to e, is governed by the same

> laws and exactly matches a structure of events $\mathscr{S}^*$, then the counterpart c^* of c in $\mathscr{S}^*$ is a cause of the counterpart e^* of e in $\mathscr{S}^*$.

There are counterexamples to this principle, on any reasonable way of spelling out what it takes for a structure of events to include all the events involved in a process connecting c to e (see Chapter 2 section 1.3.3). My account of causation sides with our intuitive judgements on these counterexamples – and thereby goes against *Intrinsicness of causation*. However, as we shall see, my account supports a restricted principle of intrinsicness. Before we can get to this, however, we first need to get a more precise understanding of the intrinsicness principle itself.

Lewis gives the intuitive motivation for the intrinsicness principle as follows:

> 'Suppose we have processes – courses of events, which may or may not be causally connected – going on in two distinct spatiotemporal regions, regions of the same or of different possible worlds. Disregarding the surroundings of the two regions, and disregarding any irrelevant events that may be occurring in either region without being part of the process in question, what goes on in the two regions is exactly alike. Suppose further that the laws of nature that govern the two regions are exactly the same. Then can it be that we have a causal process in one of the regions but not in the other? It seems not. Intuitively, whether a process going on in a region is causal depends only on the intrinsic character of the process itself, and on the relevant laws. The surroundings, and even other events in the region, are irrelevant.'[10]

What Lewis makes here is in fact a supervenience claim: namely, the claim that the obtaining of a causal relation supervenes on the laws together with the intrinsic nature of the cause, the effect, and the intermediate events that make up the process connecting the cause to the effect.

[10] Lewis (1986*e*), p. 205.

However, Lewis's statement is objectionably vague: in particular, he leaves it unspecified exactly what he means by a 'process'[11] – and thereby, he leaves it unspecified exactly what should be included in the supervenience base. Before we proceed, we therefore need a more precise statement of the intrinsicness principle. In the following, I begin by defining the relevant notion of matching. I next define what must be included in a structure of events $\mathscr{S}$ in order for $\mathscr{S}$ to include all the events that are involved in a process from c to e. And finally, I present my more precise statement of the intrinsicness principle.

To capture what it takes for one event or structure of events to match another, it will be useful to introduce the notion of an *image*, where one event or structure of events may be the image of another under a particular structure-preserving mapping. The kind of structure-preserving mapping I am interested in here is a mapping from time-slices in a world w to time-slices in a world w^* (where w and w^* may be the same or different worlds), which preserves before-after relations, etc. We may now define what it takes for one event to be the image of another as follows:

> *The image of an event e*: let M be a structure-preserving mapping from time-slices in a world w to time-slices in a world w^*.
> Then $e^* = (w^*, t^*, s^*, I^*, C^*)$ is the image of $e = (w, t, s, I, C)$ under M, iff
> a) t^* = the image of t under M,
> b) I^* = the image of I under M, and
> c) $C^* = C$.

Note that this notion of an image does not explicitly mention intrinsicness. However, it is still implicitly concerned with intrinsic character: as we have seen, complete states are intrinsic properties of time-slices of worlds (Chapter 3 section 2). By extension, classes of complete states are also intrinsic properties of time-slices of worlds. When an event $e^* = (w^*, t^*, s^*, I^*, C^*)$ is the

[11] Hall (2004*b*), p. 239; Paul and Hall (2013), p. 126.

image of an event $e = (w, t, s, I, C)$, this therefore tells us that the time-t^* slice of w^* shares certain intrinsic properties with the time-t slice of w: both time-slices are in a complete state that belongs to the class of complete states $C = C^*$. For example, both time-slices are in a complete state such that Suzy is throwing a rock at the window.

At the same time, the definition allows that the two time-slices need not be perfectly intrinsically alike: the complete state s^* that features in the realization of e^* may be different from the complete state s that features in the realization of e. Thus, it may for example be the case that the complete state that features in the realization of e is such that Suzy is throwing a rock at the window and a blackbird is singing from a tree, while the complete state that features in the realization of e^* is such that Suzy is throwing a rock at the window, and there are no birds in the tree. This is as it should be, since it would trivialise the intrinsicness principle if we required perfect match of complete states.

Since we are only concerned with structure-preserving mappings, it never happens that two different time-slices in w are mapped onto the same time-slice in w^*, and thus, it never happens that a single event in w^* is the image of two different events in w. Hence every image-event e^* has a well-defined pre-image e. Building on this, we may now define what it takes for one structure of events to be the image of another:

> *The image of a structure of events*: let M be a structure-preserving mapping from time-slices in a world w to time-slices in a world w^*.
> Then a structure $\mathscr{S}^*$ of events occurring in w^* is the image under M of a structure $\mathscr{S}$ of events occurring in w iff:
> an event e belongs to $\mathscr{S}$ just in case the image e^* of e under M belongs to $\mathscr{S}^*$.

Next, we need to answer the question of what it takes for a structure of events to include all the events that are involved in a process from c to e. Giving a fully general answer to this question is a relatively complex matter. I take up this task in Appendix A section 3.1. For our present purposes, however, we may limit ourselves to neuron diagrams – and here, the general proposal presented in Appendix A section 3.1 entails a simple and intuitively correct answer to our question:

When we are considering a neuron diagram, we may identify a time-series that includes all those times at which the neuron diagram explicitly shows what happens. We may then give a characterisation of a process based on such a time-series. Take, for example, our standard case of early preemption:

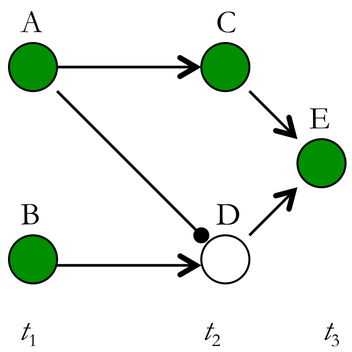


Figure 3

In this case, the time-series T: $t_1 < t_2 < t_3$ includes all those times at which the neuron diagram explicitly shows what happens. And based on this time-series, we may characterise the process from A to E as follows:

t_1: A
t_2: C
t_3: E

To include all the events that are involved in a process from c to e, a structure of events $\mathscr{S}$ must include all the events that feature in such a characterisation of a process from c to e. In the present case, for example, our structure of events needs to include the three events A, C, and E.

Furthermore, when one event stands in the right relation to another, this is sometimes so in virtue of other events. To bring this out, consider the following case of overdetermined joint causation:

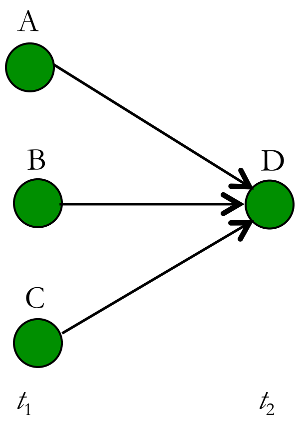


Figure 5

In this case, **D** fires if and only if it receives signals from both **A** and **B**, or from **C**. We here find that there is a process from *A* to *D*, namely:

t_1:	*A*	(together with *B*)
t_2:	*D*	

A here stands in the right relation to *D* in virtue of the fact that *A* belongs to the set {*A*, *B*}, which is minimally sufficient for *D*. Consider now the following case where **B** does not fire:

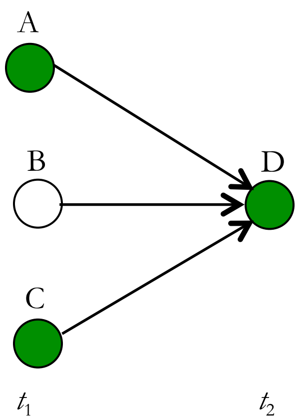


Figure 5*

In Figure 5*, *A* and *D* still occur. However, *A* is not process-connected to *D*. To have a credible version of the intrinsicness principle, we therefore need to count the events with which *A* combines – in this case, *B* – as also being involved in the process from *A* to *D*.[12]

My suggestion is that a structure of events $\mathscr{S}$ includes all the events that are involved in a process from *c* to *e* just in case it satisfies the conditions outlined above. This may be summed up as follows:

> *Including all events involved in a genuine process – applied to neuron diagrams*:
> a structure of events $\mathscr{S}$ includes all the events in a genuine process from *c* to *e* iff there is a series of events *P*, based on a time-series that includes all those times at which the neuron diagram explicitly represents what happens, such that
> a) *P* characterises a genuine process from *c* to *e*,
> b) if an event belongs to *P*, then it belongs to $\mathscr{S}$, and
> c) for any two events e_i and e_{i+1}, where e_i is followed immediately by e_{i+1} in *P*, e_i belongs to a set S_i of contemporaneous events, such that
> i) S_i is minimally sufficient for e_{i+1}, and
> ii) S_i is a subset of $\mathscr{S}$.

I suggest that we should understand the intrinsicness principle as saying that when *c* causes *e*, then any structure of events that satisfies the above definition constitutes a supervenience base for the fact *c* causes *e*. In our case of early preemption, for example, the intrinsicness principle says that the fact that *A* causes *E* supervenes on the laws together with the structure of events $\mathscr{S} = \{A, C, E\}$. And in our case of overdetermined joint causation illustrated in Figure 5, the principle says that the fact that *A* causes *D* supervenes on the laws together with the structure of events $\mathscr{S} = \{A, B, D\}$.

[12] Cf. Hall (2002), pp. 264-65; Paul and Hall (2013), p. 126.

Based on this, we may now give the following more precise statement of the principle that causation is intrinsic to a process (the asterisk indicates that I do not endorse the principle):

> **Intrinsicness of causation*: let c and e be events occurring in world w.
> Let c be a cause of e, and let $\mathscr{S}$ be a structure of events occurring in w, including all the events involved in a genuine process from c to e.
> Let $\mathscr{S}^*$ be a structure of events occurring in w^*, such that
> a) w and w^* are governed by the same laws, and
> b) $\mathscr{S}^*$ is the image of $\mathscr{S}$ under a structure-preserving mapping M.
> Then the image c^* of c under M is a cause of the image e^* of e under M.

In the following section, I will now show how this more precise understanding of the intrinsicness principle allows us to recover a restricted intrinsicness principle.

3.2 A restricted principle of intrinsicness

On my proposed account of causation, *Intrinsicness of causation* is false. However, as I prove in Appendix A section 3, process-connection *is* intrinsic to a process:

> *Intrinsicness of process-connection*: let c and e be events occurring in world w.
> Let c be process-connected to e, and let $\mathscr{S}$ be a structure of events occurring in w, including all the events involved in a genuine process from c to e.
> Let $\mathscr{S}^*$ be a structure of events occurring in w^*, such that
> a) w and w^* are governed by the same laws, and
> b) $\mathscr{S}^*$ is the image of $\mathscr{S}$ under a structure-preserving mapping M.
> Then the image c^* of c under M is *process-connected* to the image e^* of e under M.

Thus, we find that my account of causation entails the following restricted intrinsicness principle:

> *Restricted intrinsicness of causation*: let *c* and *e* be events occurring in world *w*.
> Let *c* be a cause of *e*, and let $\mathscr{S}$ be a structure of events occurring in *w*,
> including all the events involved in a genuine process from *c* to *e*.
> Let $\mathscr{S}$* be a structure of events occurring in *w**, such that
> a) *w* and *w** are governed by the same laws, and
> b) $\mathscr{S}$* is the image of $\mathscr{S}$ under a structure-preserving mapping *M*.
> Then the image *c** of *c* under *M* is a cause of the image *e** of *e* under *M*
> within a possibility horizon $\mathscr{H}$, *provided that e* security-depends on c* within* $\mathscr{H}$.

As we shall now see, this is – in spite of the restriction – a powerful principle of causal reasoning.

3.3 Applications

The principle of *Restricted intrinsicness of causation* allows us to transfer results from one case to another – in particular, it allows us to transfer results from simple environments to more complex environments. In this way, the *Restricted intrinsicness of causation* allows us to 'discover causal structure embedded within a complex system of events'[13] – without having to come to grips with the whole complex system all at once.

This makes the *Restricted intrinsicness of causation* a powerful tool in our causal reasoning. Indeed, it seems plausible that the assumption of some kind of intrinsicness principle underlies much of our scientific practice, where our method precisely is to study various phenomena under controlled conditions, and then subsequently transfer our conclusions to more complex environments.

The power of the *Restricted intrinsicness of causation* as a tool for reasoning about causation is especially clear when we are dealing with cases of redundant causation. To illustrate this, I first consider our standard case of early

[13] Paul and Hall (2013), p. 127.

preemption, and next a more complex case of redundant omission-involving causation. Consider first the following case:

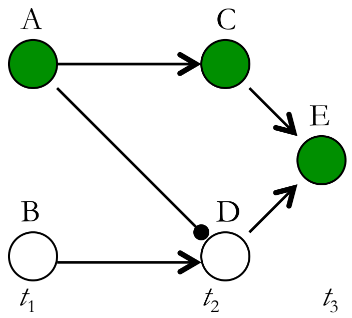


Figure 6

It is easy to see that *A* here causes *E*. Indeed, *E* depends counterfactually on *A* within any possibility horizon that treats *A* as a candidate cause. Correspondingly, we find that *A* is process-connected to *E* via the process:

t_1: *A*
t_2: *C*
t_3: *E*

By the above definition, we now find that the structure of events $\mathscr{S} = \{A, C, E\}$ includes all the events that are involved in the process from *A* to *E*.

Next, consider our standard case of early preemption:

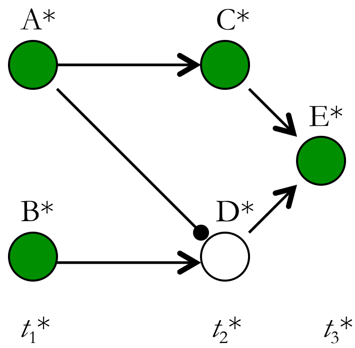


Figure 6*

It is clear that the Figure 6 and Figure 6* are governed by the same laws. Furthermore, the structure of events $\mathscr{S}^* = \{A^*, C^*, E^*\}$ is the image of the structure of events $\mathscr{S} = \{A, C, E\}$ under the structure-preserving mapping that maps t_1 to t_1*, t_2 to t_2*, etc. Finally, we have already seen that E* security-depends on A* (cf. Chapter 9 section 1.1), within the possibility horizon $\mathscr{H}$ that treats both A* and B* as candidate causes. Based on this, the *Restricted intrinsicness of causation* yields the desired result – namely, that A* causes E* within possibility horizon $\mathscr{H}$.

As a further example, let us consider how the principle applies in a case of redundant causation by omission.[14] Consider first the following simple case:

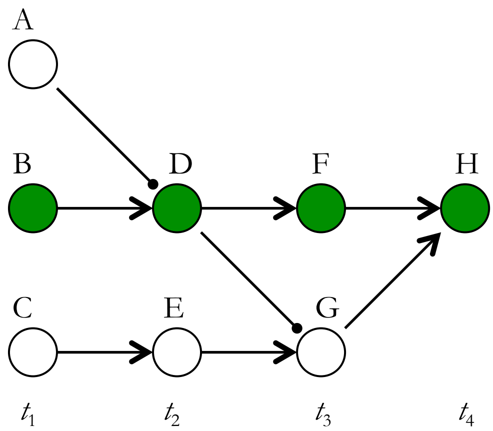


Figure 7

In this case, it is clear that $\neg A$ is a cause of H. Indeed, H depends counterfactually on $\neg A$ within any possibility horizon that treats $\neg A$ as a candidate cause. Correspondingly, we find that $\neg A$ is process-connected to H via the following process:

t_1: $\neg A$ (together with B)
t_2: D
t_3: F
t_4: H

[14] Note that the *Restricted intrinsicness of causation* applies equally well to omission-involving causation, in keeping with my general suggestion that there is no deep metaphysical distinction between omissions and absences and so-called positive events. By contrast, the intrinsicness principle proposed in Hall (2004*b*) and Paul and Hall (2013) does not apply to omission-involving causation (see Hall (2004*b*) and Paul and Hall (2013), pp. 196-97).

By the above definition, we thus find that the structure of events $\mathscr{S} = \{\neg A, B, D, F, H\}$ includes all the events involved in the process from $\neg A$ to H.

Next, consider the following case of redundant causation by omission:

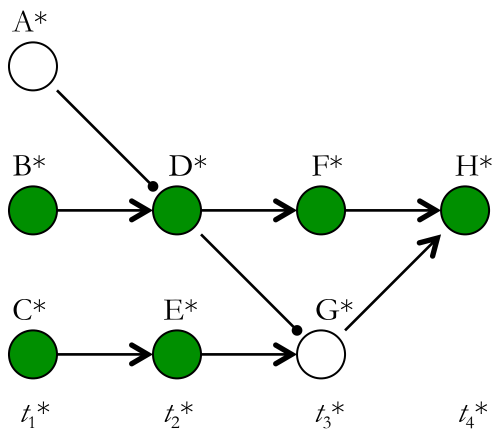


Figure 7*

It is clear that Figure 7 and Figure 7* are governed by the same laws. Furthermore, the structure of events $\mathscr{S}^* = \{\neg A^*, B^*, D^*, F^*, H^*\}$ is the image of the structure of events $\mathscr{S} = \{\neg A, B, D, F, H\}$, under the structure-preserving mapping that maps t_1 to t_1^*, t_2 to t_2^*, etc. And finally, it is easy to verify that H^* security-depends on $\neg A^*$ within any possibility horizon that treats $\neg A^*$ and C^* as candidate causes. On this basis, the *Restricted intrinsicness of causation* delivers the desired result – namely, that $\neg A^*$ is a cause of H^* within any such possibility horizon $\mathscr{H}$.

As these cases illustrate, the *Restricted intrinsicness of causation* allows us to easily handle cases of redundant causation by considering the relevant causal process in a simpler environment, where there is no redundancy, and then transferring our results to the case at hand. Although we do not *need* the intrinsicness principle – any result that can be obtained by applying the *Restricted intrinsicness of causation* can also be obtained by applying the two conditions of process-connection and security-dependence directly – the principle thus provides a useful shortcut in our causal reasoning. It is, however, important to be aware of the limitations of the principle. I discuss those limitations in the following section.

3.4 Limitations: idealisations and real-life cases

The hope that motivated Lewis's original suggestion that causation is intrinsic to a process, as well as Hall's later development of the idea, is that some kind of intrinsicness principle can provide a solution to the problem of late preemption.[15]

In support of this idea, Paul and Hall suggest that it would provide an attractive explanation of how we arrive at our intuitive judgements about cases of late preemption:

> 'consider a plausible claim about late preemption cases: we can arrive at a case of late preemption by beginning with a case of perfectly ordinary, garden variety causation involving a structure of events S, and then adding details to it that concern matters extrinsic to S. […] In this way, we can arrive at the case of Billy, Suzy, and the [window] by starting with a situation in which Suzy alone throws a rock at the [window], and then adding extrinsic details (Billy and his throw), and so on. If, as our intrinsicness thesis states, such extrinsic changes make no difference to the causal structure of the process we begin with, then it is no surprise that cases of late preemption evoke such clear and firm intuitive judgements: we "see" embedded in them, as it were, perfectly ordinary cases of causation.'[16]

There are several problems with relying on intrinsicness to handle late preemption, not least that there are convincing counterexamples to the unrestricted intrinsicness principle.[17] As we have seen, however, my account of

---

[15] See Hall (2002) and (2004*b*), pp. 258-65; Lewis (1986*e*), pp. 205-7.

[16] Paul and Hall (2013), p. 128.

[17] In recognition of these counterexamples, Lewis retracts his endorsement of the intrinsicness principle (Lewis (2004*a*), pp. 83-85, while Hall limits the principle to one of his two concepts

causation entails a *restricted* intrinsicness principle. To what extent can the *Restricted intrinsicness of causation* fulfil Lewis's and Hall's ambitions of using some form of intrinsicness principle to handle cases of late preemption?

Of course, the *Restricted intrinsicness of causation* does not *need* to fulfil these ambitions: we already have a way to distinguish genuine causes from preempted backups – namely by applying the condition of process-connection directly, as illustrated in Chapter 7. However, it will nevertheless be illuminating to see how the *Restricted intrinsicness of causation* relates to cases of late preemption: as we shall see, the *Restricted intrinsicness of causation* does fulfil Lewis's and Hall's ambitions in simple and idealised cases of late preemption, but fails to do so in complex real-life cases. This result in turn points to an interesting observation concerning the role of idealisation.

Let us begin by seeing how the *Restricted intrinsicness of causation* applies to a simple case of late preemption:

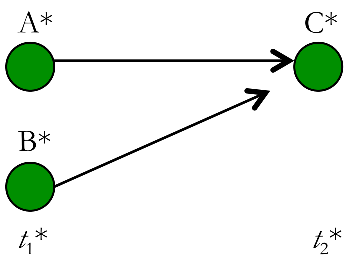


Figure 8*

Suppose that the speed of the stimulatory signal from **A*** to **C*** is independent of whether or not **B*** fires. Suppose further that the stimulatory signal from **A*** reaches **C*** exactly at time t_2*, whereupon **C*** immediately fires. By contrast, the stimulatory signal from **B*** reaches **C*** at some slightly later time – say, at time t_3*. Suppose finally that the effect we are interested in is – against our usual practices when dealing with neuron diagrams – a temporally robust event, namely the event *C** based on the interval *I** = [t_2*, t_3*], and the class of states such that **C*** fires. Based on this, we find that the case has the standard

of causation, namely the concept of production (Hall (2004*b*)). For discussion, see also Paul (1998*b*).

characteristics of a late preemption case: intuitively, *A** is a cause of *C**, while *B** is not – but if *A** had not occurred, *C** would still have occurred because of *B** (it would simply have occurred slightly later – namely, at t_3* instead of t_2*).

To see how the *Restricted intrinsicness of causation* allows us to handle this case, let us begin by considering the following simpler case:

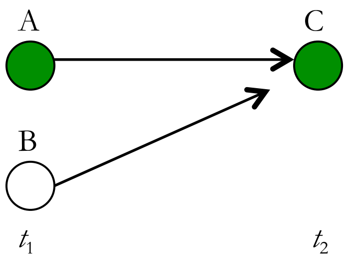


Figure 8

In this simple case, it is clear that *A* causes *C*. Indeed, *C* depends counterfactually on *A* within any possibility horizon that treats *A* as a candidate cause. Correspondingly, we find that there is a genuine process from *A* to *C*, namely:

t_1: *A*

t_2: *C*

Thus, the structure of events $\mathscr{S} = \{A, C\}$ includes all the events involved in the process from *A* to *C*.

Let us now compare this case with our case of late preemption:

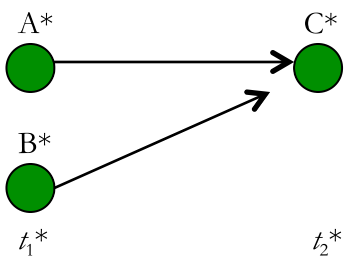


Figure 8*

It is clear that Figure 8 and Figure 8* are governed by the same laws. It is also clear that the structure of events $\mathscr{S}^* = \{A^*, C^*\}$ is the image of the structure of events $\mathscr{S} = \{A, C\}$ under the structure-preserving mapping that maps time t_1 to t_1^*, etc. Finally, it is easily seen that C^* security-depends on A^* within the possibility horizon $\mathscr{H}$ that treats both A^* and B^* as candidate causes. Thus, the *Restricted intrinsicness of causation* allows us to conclude that A^* causes C^* within possibility horizon $\mathscr{H}$.

So far, then, the *Restricted intrinsicness of causation* satisfies Lewis's and Hall's ambitions: it provides a straightforward way to transfer conclusions about what causes what in simple environments to cases of late preemption, and it allows us to explain our intuitions about such cases as a result of how 'we "see" embedded in them […] perfectly ordinary cases of causation.'[18]

Once we attempt to apply the *Restricted intrinsicness of causation* to real-life cases, such as the case of Suzy, Billy, and the window, however, the limitations of the principle become apparent. This should hardly be a surprise – the difficulty in applying any kind of intrinsicness principle to real life cases such as this one has long been recognised by Lewis and Hall.[19] The problem is that the process from Suzy's throw to the window-shattering in the simple case where Billy is absent does not *perfectly* match the process from Suzy's throw to the window-shattering in the case where Billy is there and throws a rock too:

> 'because of Billy's absence, there will, of nomological necessity, be ever so slight intrinsic differences in the structure of events consisting of the flight of Suzy's rock towards the [window].'[20]

This places a strain on the formulation of the intrinsicness principle. As Paul and Hall write:

[18] Paul and Hall (2013), p. 128.

[19] See e.g. Hall (2002), p. 287 and (2004*b*), p. 259; Lewis (2004*a*), p. 82.

[20] Paul and Hall (2013), p. 131. See also Hall (2002), p. 287; Hall (2004*b*), p. 259; Lewis (2004*a*), p. 82.

> 'an application of [the intrinsicness thesis] to the problem posed by late preemption requires some delicacy when laying out the relevant notion of "matching": perfect match of intrinsic character makes for a more readily defensible thesis, but at the cost of rendering that thesis useless as a cure for the late preemption disease.'[21]

To address this problem, Hall has an interesting discussion of how we might distinguish between relevant and irrelevant respects of similarity.[22] On my account of causation, however, we do not *need* any kind of intrinsicness principle to handle problems of late preemption: we can handle them by applying my account directly. I am therefore content to only satisfy part of Lewis's and Hall's ambitions: as I set out in more detail in Appendix A section 3.1, a sufficiently comprehensive structure of events must include all those events that influence the timing of the events in our target process. As a result, the *Restricted intrinsicness of causation* can only be used when reasoning about simple cases of late preemption, such as the one illustrated in Figure 8* above; it simply does not apply in complex real life cases such as the case of Suzy, Billy, and the window.

However, this does not mean that the *Restricted intrinsicness of causation* has no relevance for understanding how we think about such real life cases: at least in our scientific practice – and plausibly in our everyday reasoning too – we do in fact make use of exactly this kind of intrinsicness principle in thinking about complex real life cases. To see how, we first need to look more closely at the role of *idealisations*.

Making idealisations is an important part of our scientific practice. When calculating the trajectory of a rock, for example, we usually focus only on significant contributions such as the rock's momentum, the gravitational field from the earth, etc. – while disregarding the comparatively much smaller

[21] Paul and Hall (2013), p. 196.

[22] See Hall (2002), pp. 286-290.

contributions from air resistance, air currents, gravitational fields from small objects in the vicinity, etc. This yields a much cleaner and more manageable – though not entirely truthful – picture of the situation at hand.

If we allow ourselves to make such idealisations, we *can* apply the *Restricted intrinsicness of causation* – we can apply it to the *idealised* version of the case at hand. We may, for example, consider an idealised version of our late preemption case with Suzy, Billy, and the window, where we disregard small contributions such as air resistance, air currents, gravitational fields from small objects in the vicinity etc. When we are considering such an idealised version of the case, the process from Suzy's throw to the window-shattering in the case where Billy is absent *does* perfectly match the process from Suzy's throw to the window-shattering in the case where Billy is there and throws a rock too – and this allows us to apply the *Restricted intrinsicness of causation* in the idealised case.

It is important to note, however, that bringing in idealisations like this does not allow the *Restricted intrinsicness of causation* to *solve* the problem of late preemption: for that, we still need the direct application of my account of causation, and in particular the condition of process-connection. The reason for this is that the idealisations are *prior* to the application of the *Restricted intrinsicness of causation*; that principle in itself does nothing to tell us which idealisations are permissible and which idealisations will lead us astray. To distinguish the good idealisations from the bad, we thus need to fall back on the results delivered by applying my account of causation directly.[23]

4. Conclusion

What I hope to have shown in this chapter is that, while my proposed account of causation cannot accommodate the three intuitive principles of causal

[23] This has clear parallels in the case of science, where the art of making good approximations and idealisations – namely, approximations and idealisations that make the cases manageable, while delivering an appropriate degree of accuracy in the final results – is in the same way something that comes *before* the application of the theories themselves, and is constrained instead by experiments and experience with more detailed applications of the theories.

inference – concerning the sufficiency of counterfactual dependence, transitivity, and intrinsicness – in their unrestricted form, it can accommodate restricted versions of all three principles. These restricted versions are compatible with our intuitive verdicts on counterexamples to the unrestricted principles. But at the same time, they allow us to preserve our core practices of causal reasoning. In this way, my account of causation shows charity to use not just in accommodating our intuitive judgements about specific cases, but also by entailing that our core practices of causal reasoning are legitimate.

12

Conclusion

I started with the question: what is causation? I am now in a position to present my answer in the form of the following biconditional:

> *Causation*: *c* is a cause of *e* within a possibility horizon $\mathscr{H}$ iff
>
> a) *c* is process-connected to *e*, and
>
> b) *e* security-depends on *c* within $\mathscr{H}$.

This biconditional captures the dual nature of causation: a cause must be connected to its effect via a genuine process, and it must make a difference to its effect.

The first condition – namely, the condition of *process-connection* – captures the sense in which a cause must be connected to its effect via a genuine process. As we have seen, this condition allows my account to separate causation from mere correlation, distinguish genuine causes from preempted backups, and capture the sense in which a cause must be at the right level of detail relative to its effect (Chapter 7).

The second condition – namely, the condition of *security-dependence* – captures the sense in which a cause must make a difference to its effect. As we have seen, this condition allows my account to deliver intuitively correct verdicts on the counterexamples to the transitivity and intrinsicness of causation, resolve the problem of profligate omissions, accommodate structurally isomorphic but causally different cases, and handle contrastive causal claims (Chapter 9 and 10).

Recognising the dual nature of causation also allows my account to support restricted versions of the three intuitive principles concerning the sufficiency of counterfactual dependence for causation, and the transitivity and intrinsicness of causation. The reason for this is that each of the three intuitive principles is satisfied by one, but not the other, of the two relations of process-connection and security-dependence: counterfactual dependence within a possibility horizon is sufficient for security-dependence within that same possibility horizon, but is not sufficient for process-connection; and process-connection is transitive and intrinsic to a process, while security-dependence is not. As a result, we find that the following three restricted principles are logical consequences of my account of causation (Chapter 11):

> *Restricted sufficiency of counterfactual dependence for causation*:
> if *e* depends counterfactually on *c* within a possibility horizon $\mathscr{H}$,
> such that $\mathscr{H}$ is discriminating at *c*'s time of occurrence,
> then *c* or a more or less fragile version of *c* is a cause of *e* within $\mathscr{H}$.
>
> *Restricted transitivity of causation*: if there is a set of events, $\{d_1, d_2, \ldots, d_n\}$, such that *c* causes d_1, d_1 causes d_2, . . . , and d_n causes *e*, then *c* causes *e* within a possibility horizon $\mathscr{H}$, *provided that e security-depends on c within* $\mathscr{H}$.
>
> *Restricted intrinsicness of causation*: let *c* and *e* be events occurring in world *w*.
> Let *c* be a cause of *e*, and let $\mathscr{S}$ be a structure of events occurring in *w*,
> including all the events involved in a genuine process from *c* to *e*.
> Let $\mathscr{S}$* be a structure of events occurring in *w**, such that
> a) *w* and *w** are governed by the same laws, and
> b) $\mathscr{S}$* is the image of $\mathscr{S}$ under a structure-preserving mapping *M*.
> Then the image *c** of *c* under *M* is a cause of the image *e** of *e* under *M* within a possibility horizon $\mathscr{H}$, *provided that e* security-depends on c* within* $\mathscr{H}$.

Finally, treating causation as a ternary relation with a possibility horizon as its third relatum allows my account to accommodate the context-sensitivity of our causal claims: metaphysically, causation is a ternary relation – *c* causes *e* within possibility horizon $\mathscr{H}$. It is an entirely context-independent matter whether this ternary relation holds. Usually, however, the surface form of our causal claims is binary – as in '*c* causes *e*' – and we rely on context to supply the relevant possibility horizon. And this accounts for the fact that the truth-value of our binary causal claims depends on context: one context may select a possibility horizon $\mathscr{H}_1$, such that it is true that *c* causes *e* within $\mathscr{H}_1$; a different context may select a different possibility horizon $\mathscr{H}_2$, such that it is false that *c* causes *e* within $\mathscr{H}_2$.

It is, however, important to note that the context-sensitivity of causal claims is limited: process-connection is a binary relation between instantaneous events. Since process-connection is a necessary condition for causation, this means that whenever an event *c* is *not* process-connected to an event *e*, the causal claim that '*c* does not cause *e*' is true independently of our choice of possibility horizon. For example, if *c* is merely a preempted backup in relation to *e*, it is true independently of context that *c* does not cause *e*. Context only plays a role in the selection among those events that are process-connected to the effect: security-dependence is a ternary relation between two instantaneous events and a possibility horizon, and thus, we may find that *e* security-depends on *c* within one possibility horizon $\mathscr{H}_1$, but does not security-depend on *c* within a different possibility horizon $\mathscr{H}_2$.

How well does this account of causation satisfy the criteria of success set out in Chapter 1?

First, my account provides an ontological reduction of causation: my account is based on four basic ingredients – complete states, laws of nature understood as rules specifying the temporal evolution of complete states, relations of overall similarity between complete states, and the space of metaphysically possible worlds. On the basis of these four ingredients, I define the two kinds of causal relata – namely, *instantaneous events* (Chapter 4) and *possibility horizons* (Chapter 5) – and the two relations of *process-connection* (Chapter 6) and *security-dependence* (Chapter 8). These are the only ingredients that go into my definition of the ternary relation of causation within a possibility horizon. Thus, they are the only ingredients that go into my account of what causation is out in the world.

My account of how context selects a possibility horizon is not, strictly speaking, part of my account of what causation is out in the world: it simply captures how we take an interest in some ways of completing the statement '*c* causes *e*', but not others. To verify that my account of causation yields intuitively correct results when applied to our ordinary causal statements, however, it is indispensable to have an account of how we should select the third relatum when we are evaluating a binary causal claim of the form '*c* causes *e*'. To capture this, I bring in a further ingredient – namely, the descriptive fact that we accept certain extrapolations as appropriate and reject others. This notion of extrapolation is non-causal, and thus we can capture – in a non-causal way – how context selects the relevant possibility horizon.

On this basis, I believe that my proposed account succeeds in providing an ontological reduction of causation.

Second, I will now argue that my account achieves a high degree of eligibility and charity to use:

In developing my account of causation, I have been guided by considerations of eligibility in two important ways. Firstly, there is a close connection between eligibility and ontological reduction: an ontological reduction shows how the causal relation is built up from fundamental features of the world – complete states, laws of nature, etc. And other things being equal, a relation that has a relatively simple definition in terms of these fundamental features of the world is more eligible than a relation that does not. The fact that I propose an ontological reduction therefore speaks in favour of the eligibility of my proposed candidate meaning of 'cause'.

The second way in which I have been guided by considerations of eligibility is in my search for relations with beautiful formal properties. I believe that it speaks in favour of the eligibility of the relation of process-connection that it is transitive and intrinsic to a process. I similarly believe that it speaks in favour of the eligibility of the relation of security-dependence that counterfactual dependence within a possibility horizon entails security-dependence within that same possibility horizon. And by extension, I believe that it speaks in favour of the eligibility of my proposed candidate meaning of 'cause' that it can be defined in terms of these two relations.

Furthermore, I believe that my proposed account of causation achieves a high degree of charity to use. Indeed, as far as I am aware, the only cases where my proposed account goes against our intuitive judgements are cases where those intuitive judgements come into conflict with each other: in counterexamples to the sufficiency of counterfactual dependence, and to the transitivity and intrinsicness of causation, my account sides with our intuitive judgements about specific cases, against the intuitive principles in their unrestricted form. On the other hand, there are certain more complex cases where my account goes against our intuitive judgements about specific cases, because our intuitive judgements about these cases conflict with the independently plausible *restricted* principles of sufficiency of counterfactual dependence, transitivity, and intrinsicness.

By striking this balance between our intuitive judgements about specific cases and our intuitive acceptance of the three general principles concerning the sufficiency of counterfactual dependence for causation, and the transitivity and intrinsicness of causation, I believe that my proposed account achieves a high degree of charity to use: it yields intuitively correct verdicts on almost all the specific cases we have considered (for an overview, see Appendix B), and at the same time, it legitimises our core practices of appealing to counterfactual dependence, transitivity, and intrinsicness in our causal reasoning.

References

Achinstein, Peter. 1975. 'Causation, Transparency, and Emphasis.' *Canadian Journal of Philosophy* 5: 1-23.

Armstrong, David M. 1978*a*. *Nominalism and Realism*, volume 1 of *Universals and Scientific Realism*. Cambridge: Cambridge University Press.

Armstrong, David M. 1978*b*. *A Theory of Universals*, volume 2 of *Universals and Scientific Realism*. Cambridge: Cambridge University Press.

Armstrong, David. 2004. 'Going Through the Open Door Again: Counterfactual vs. Singularist Theories of Causation.' In Collins et al. 2004*a*, 445-47.

Beebee, Helen. 2004. 'Causing and Nothingness.' In Collins et al. 2004*a*, 291-308.

Bernstein, Sara. 2014. 'Two Problems for Proportionality about Omissions.' *Dialectica* 68: 429-41.

Bernstein, Sara. 2015. 'A Closer Look at Trumping.' *Acta Analytica* 30: 41-57.

Blanchard, Thomas and Jonathan Schaffer. *Forthcoming*. 'Cause without Default'. To appear in Beebee, Hitchcock, and Price, eds. *Making a Difference*.

Collins, John. 2004. 'Preemptive Prevention.' In Collins et al. 2004*a*, 107-117.

Collins, John, Ned Hall, and L. A. Paul, eds. 2004*a*. *Causation and Counterfactuals*. Cambridge, MA: MIT Press.

Collins, John, Ned Hall, and L. A. Paul. 2004*b*. 'Introduction'. In Collins et al. 2004*a*.

Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Dowe, Phil. 2010. 'Proportionality and Omissions.' *Analysis* 70: 446-51.

Ehring, Douglas. 2009. 'Causal Relata.' In H. Beebee, C. Hitchcock and P. Menzies, eds. *The Oxford Handbook of Causation*. Oxford: Oxford University Press, pp. 387-413.

Fair, David. 1979. 'Causation and the Flow of Energy.' *Erkenntnis* 14: 219-50.

Forrest, Peter. 1995. 'Is Space-Time Discrete or Continuous? An Empirical

Question.' *Synthese* 103: 327-54.

Gardner, Martin. 1970. 'Mathematical Games: the Fantastic Combinations of John Conway's New Solitaire Game "Life".' *Scientific American* 223: 120-23.

Goodman, Nelson. 1954. *Fact, Fiction, and Forecast.* London: the Athlone Press.

Hall, Ned. 2002. 'The Intrinsic Character of Causation.' In Dean Zimmerman, ed. *Oxford Studies in Metaphysics*, volume 1. Oxford: Oxford University Press.

Hall, Ned and L. A. Paul. 2003. 'Causation and Pre-emption.' In Peter Clark and Katherine Hawley, eds. *Philosophy of Science Today*. Oxford: Clarendon Press.

Hall, Ned. 2004*a*. 'Causation and the Price of Transitivity.' In Collins et al. 2004*a*, 181-203.

Hall, Ned. 2004*b*. 'Two Concepts of Causation.' In Collins et al. 2004*a*, 225-76.

Hall, Ned. 2007*a*. 'Structural Equations and Causation.' *Philosophical Studies* 132: 109-136.

Hall, Ned. 2007*b*. 'Structural Equations and Causation', extended version. http://nrs.harvard.edu/urn-3:HUL.InstRepos:3710361.

Halpern, J. and J. Pearl. 2005. 'Causes and Explanations: A Structural Model Approach.' *British Journal for the Philosophy of Science* 56: 889-911.

Halpern, Joseph and Christopher Hitchcock. 2015. 'Graded Causation and Defaults'. *British Journal for the Philosophy of Science* 66: 413-57.

Hart, H. L. A. and A. M. Honoré. 1985. *Causation in the Law*. Oxford: Clarendon Press.

Hawley, Katherine. 2004. *How Things Persist.* Oxford: Clarendon Press.

Hitchcock, Christopher. 1996. 'The Role of Contrast in Causal and Explanatory Claims.' *Synthese* 107: 395-419.

Hitchcock, Christopher. 2001. 'The Intransitivity of Causation Revealed in Equations and Graphs,' *Journal of Philosophy* 98: 273-99.

Hitchcock, Christopher. 2004. 'Do All and Only Causes Raise the Probabilities of Effects?' In Collins et al. 2004*a*, 403-417.

Hitchcock, Christopher. 2007. 'Prevention, Preemption, and the Principle of

Sufficient Reason.' *The Philosophical Review* 116: 495-532.

Hitchcock, C. 2009. 'Structural Equations and Causation: Six Counterexamples.' *Philosophical Studies* 144: 391-401.

Hitchcock, Christopher and J. Knobe. 2009. 'Cause and Norm.' *Journal of Philosophy* 106: 587-612.

Hume, David. 2007 (first published in 1739). *A Treatise of Human Nature*. Oxford: Clarendon Press.

Kaiserman, Alex. Forthcoming. 'Causes and Counterparts'. *Australasian Journal of Philosophy*.

Kim, Jaegwon. 1973. 'Causation, Nomic Subsumption, and the Concept of Event.' *Journal of Philosophy* 70: 217-36.

Kim, Jaegwon. 1976. 'Events as Property Exemplifications.' In Brand, Myles and Douglas Walton, eds. *Action Theory*, Dordrecht: D. Reidel Publishing Company, 159-77.

Kutach, Douglas. 2007. 'The Physical Foundations of Causation.' In Price et al. 2007, 327-350.

Kvart, Igal. 1991. 'Transitivity and Preemption of Causal Relevance.' *Philosophical Studies* 64: 125-60.

Langton, Rae and David Lewis. 1998. 'Defining 'Intrinsic'.' *Philosophy and Phenomenological Research* 58: 333-45.

Lee, Jig-chuen. 1988. 'The Nontransitivity of Causation.' *American Philosophical Quarterly* 25: 87-94.

Le Poidevin, Robin. 1990. 'Relationism and Temporal Topology: Physics or Metaphysics?' *The Philosophical Quarterly* 40: 419-32.

Lewis, David. 1973. *Counterfactuals*. Cambridge: Harvard University Press.

Lewis, David. 1983. 'New Work for a Theory of Universals.' *Australasian Journal of Philosophy* 61: 343-77.

Lewis, David. 1984. 'Putnam's Paradox.' *Australasian Journal of Philosophy* 62: 221-36.

Lewis, David. 1986*a*. *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, David. 1986*b*. *Philosophical Papers*, Vol. II. Oxford: Oxford University Press.

Lewis, David. 1986*c*. 'Counterfactual Dependence and Time's Arrow'. In Lewis 1986*b*, 32-66.

Lewis, David. 1986*d*. 'Causation'. In Lewis 1986*b*, 159-171.

Lewis, David. 1986*e*. 'Postscripts to 'Causation'.' In Lewis 1986*b*, 172-213.

Lewis, David. 1986*f*. 'Causal Explanation.' In Lewis 1986*b*, 214-240.

Lewis, David. 1986*g*. 'Events.' In Lewis 1986*b*, 241-69.

Lewis, David. 2004*a*. 'Causation as Influence.' In Collins et al. 2004*a*, 75-106.

Lewis, David. 2004*b*. 'Void and Object.' In Collins et al. 2004*a*, 277-290.

Mackie, J. L. 1974. *The Cement of the Universe*. Oxford: Oxford University Press.

Mackie, P. 1992. 'Causing, Delaying, and Hastening: Do Rains Cause Fires?' *Mind* 101: 483-500.

Marshall, Dan and Josh Parsons. 2001. 'Langton and Lewis on 'Intrinsic'.' *Philosophy and Phenomenological Research* 63: 347-51.

Maslen, Cei. 2004. 'Causes, Contrasts, and the Nontransitivity of Causation.' In Collins et al. 2004*a*, 341-57.

Maudlin, Tim. 2004. 'Causation, Counterfactuals, and the Third Factor.' In Collins et al. 2004*a*, 419-443.

Maudlin, Tim. 2007*a*. *The Metaphysics Within Physics*. Oxford: Oxford University Press.

Maudlin, Tim. 2007*b*. 'A Modest Proposal Concerning Laws, Counterfactuals, and Explanations.' In Maudlin 2007*a*.

McDermott, Michael. 1995. 'Redundant Causation.' *British Journal for the Philosophy of Science* 46: 523-544.

McDonnell, Neil. Forthcoming. 'Transitivity and Proportionality in Causation.' *Synthese*.

McGrath, Sarah. 2005. 'Causation by Omission: A Dilemma.' *Philosophical Studies* 123: 125-48.

Mellor, Hugh. 1995. *The Facts of Causation*. London: Routledge.

Menzies Peter. 2004. 'Difference-making in Context.' In Collins et al. 2004*a*, 139-180.

Mill, John Stuart. 1973 (first published 1843). *A System of Logic, Ratiocinative and Inductive*. London: Routledge & Kegan Paul.

Newton-Smith, W. H. 1980. *The Structure of Time*. London: Routledge & Kegan Paul.

Northcott, Robert. 2008. 'Causation and Contrast Classes.' *Philosophical Studies* 139: 111-23.

Paul, L. A. 1998*a*. 'Keeping Track of the Time: An Emended Counterfactual Analysis of Causation.' *Analysis* 58: 191-8.

Paul, L. A. 1998*b*. 'Problems with Late Preemption.' *Analysis* 58: 48-53.

Paul, L. A. 2004*a*. 'Aspect Causation.' In Collins et al. 2004*a*, 205-24.

Paul, L. A. and Ned Hall. 2013. *Causation: A User's Guide*. Oxford: Oxford University Press.

Price, Huw and Richard Corry. 2007. *Causation, Physics, and the Constitution of Reality*. Oxford: Clarendon Press.

Ramachandran, Murali. 1997. 'A Counterfactual Analysis of Causation.' *Mind* 106: 263-77.

Salmon, Wesley. 1994. 'Causation Without Counterfactuals.' *Philosophy of Science* 61: 297-312.

Sartorio, Carolina. 2005. 'Causes as Difference-Makers.' *Philosophical Studies* 123: 71-96.

Sartorio, Carolina. 2006. 'On Causing Something to Happen in a Certain Way without Causing It to Happen.' *Philosophical Studies* 129: 119-136.

Sartorio, Carolina. 2010. 'The Prince of Wales Problem for Counterfactual Theories of Causation.' In Allan Hazlett, ed. *New Waves in Metaphysics*. Basingstoke: Palgrave Macmillan.

Sartorio, Carolina. 2016. *Causation and Free Will*. Oxford: Oxford University Press.

Schaffer, Jonathan. 2000*a*. 'Causation by Disconnection.' *Philosophy of Science* 67:

285-300.

Schaffer, Jonathan. 2003. 'Overdetermining Causes.' *Philosophical Studies* 114: 23-45.

Schaffer, Jonathan. 2004. 'Trumping Preemption.' *Journal of Philosophy* 97: 165-81. Reprinted in Collins et al. 2004*a*, 59-73.

Schaffer, Jonathan. 2005. 'Contrastive Causation.' *The Philosophical Review* 114: 297-328.

Schaffer, Jonathan. 2010. 'Contrastive Causation in the Law.' *Legal Theory* 16: 259-97.

Schaffer, Jonathan. 2012. 'Causal Contextualism,' in M. Blaauw, ed. *Contrastivism in Philosophy*. New York: Routledge, 35-63.

Schaffer, Jonathan. 2016. 'The Metaphysics of Causation.' *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/

Shapiro, Larry and Elliott Sober. 2012. 'Against Proportionality.' *Analysis* 72: 89-93.

Sider, Theodore. 2001. 'Maximality and Intrinsic Properties.' *Philosophy and Phenomenological Research* 63: 357-64.

Sider, Theodore. 2003. 'Maximality and Microphysical Supervenience.' *Philosophy and Phenomenological Research* 66: 139-49.

Sider, Theodore. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.

Steglich-Petersen, Asbjørn. 2012. 'Against the Contrastive Account of Singular Causation.' *British Journal of Philosophy of Science* 63: 115-143.

Stein, Nathanael. 2016. 'Causes and Categories.' *Noûs* 50: 465-89.

Strawson, P. F. 1985. 'Causality and Explanation.' In Vermazen, Bruce and Merrill Hintikka, eds. *Essays on Davidson: Actions and Events*, Oxford: Clarendon Press, 115–36.

Thomson, Judith. 2003. 'Causation: Omissions.' *Philosophy and Phenomenological Research* 66: 81-103.

Tooley, Michael. 1990. 'Causation: Reductionism versus Realism.' *Philosophy and Phenomenological Research* 50, Supplement: 215-36.

Traynor, Michael. 2014. 'Actual Time and Possible Change: A Problem for Modal Arguments for Temporal Parts.' *Thought* 2: 180-89.

van Inwagen, Peter. 1990. *Material Beings*. Ithaca: Cornell University Press.

Weatherson, Brian. 2001. 'Intrinsic Properties and Combinatorial Principles.' *Philosophy and Phenomenological Research* 63: 365-80.

Weatherson, Brian. 2003. 'What Good are Counterexamples?' *Philosophical Studies* 115: 1-31.

Weslake, Brad. 2013. 'Proportionality, Contrast and Explanation.' *Australasian Journal of Philosophy* 91: 785-97.

Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.

Williamson, Timothy. 2016. 'Modal Science.' *Canadian Journal of Philosophy* 46: 453-92.

Woodward, James. 1984. 'A Theory of Singular Causal Explanation.' *Erkenntnis* 21: 231–62.

Woodward, James. 2003. *Making Things Happen*: *A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, James. 2006. 'Sensitive and Insensitive causation.' *Philosophical Review* 115: 1-50.

Woodward, James. 2007. 'Causation with a Human Face.' In Price et al. 2007, 66-105.

Yablo, Stephen. 1992*a*. 'Mental Causation.' *Philosophical Review* 101: 245-80.

Yablo, Stephen. 1992*b*. 'Cause and Essence.' *Synthese* 93: 403-39.

Yablo, Stephen. 2004. 'Advertisement for a Sketch of an Outline of a Prototheory of Causation.' In Collins et al. 2004*a*, 119-37.

APPENDIX

A

Proofs

In this Appendix, I prove four important results concerning the relations of process-connection and security-dependence. First, I show that process-connection satisfies a restricted principle of sufficiency of counterfactual dependence (section 1). Second, I show that process-connection is a transitive relation (section 2). Third, I show that process-connection is intrinsic to a process (section 3). And finally, I show that counterfactual dependence within a possibility horizon entails security-dependence within that same possibility horizon (section 4).

1. The restricted sufficiency of counterfactual dependence for process-connection

In this section, I show that although counterfactual dependence is not sufficient for process-connection (cf. Chapter 7 section 4), we can recover a restricted principle of sufficiency of counterfactual dependence for process-connection (section 1.1). My proof rests on a particular assumption, which I argue is satisfied in all ordinary cases (section 1.2).

1.1 The proof

My aim in this section is to show that the following restricted principle holds:

> *Restricted sufficiency of counterfactual dependence for process-connection*:
> if *e* depends counterfactually on *c* within a possibility horizon $\mathscr{H}$,
> such that $\mathscr{H}$ is discriminating at *c*'s time of occurrence,
> then *c* or a more or less fragile version of *c* is process-connected to *e*.

For ease of reference, I repeat the definition of what it takes for a possibility horizon to be discriminating (cf. Chapter 11 section 1):

> *Discriminating*: a possibility horizon $\mathscr{H}$ is *discriminating* at a time *t* iff for any event *c* occurring in @ at *t*, it is the case that either
> a) *c* occurs in every world in $\mathscr{H}$, or
> b) *c* and more or less fragile versions of *c* are the only events that occur in @ at *t*, but do not occur in the closest-at-*t* not-*c*-world(s) in $\mathscr{H}$.

Note that this definition is implicitly relativized to the contextually relevant restriction on the domain of instantaneous events, since the definition requires us to quantify over instantaneous events. Note further that the definition is trivially satisfied when we restrict our domain of instantaneous events to contain only maximally fragile instantaneous events. When we include further instantaneous events within our domain of quantification, the definition becomes correspondingly more demanding.

My proof of *Restricted sufficiency of counterfactual dependence for process-connection* will rely on the following assumption:

> *Assumption of sufficiency*: let *e* be an instantaneous event that occurs in @ at *t´*, let *t* be some time strictly earlier than *t´*, and let $\mathscr{T}$ be the set of all those events that occur in @ at *t* and are process-connected to *e*. Then $\mathscr{T}$ is sufficient for *e*.

In section 1.2, I argue that this assumption is satisfied in all ordinary cases, including all neuron diagrams. In the following, I will now prove the *Restricted sufficiency of counterfactual dependence for process-connection*, given this assumption:

Suppose that *c* occurs at time *t*, and that *e* depends counterfactually on *c* within a possibility horizon $\mathscr{H}$, i.e. *e* occurs in @ (at some time strictly later than *t*), and does not occur in the closest-at-*t* not-*c*-world(s) in $\mathscr{H}$. Assuming that $\mathscr{H}$ is discriminating at time *t*, we know that *c* and more or less fragile versions of *c* are the only events that occur in @ at *t*, but do not occur in the closest-at-*t* not-*c*-world(s) in $\mathscr{H}$.

We want to prove that *c* or a more or less fragile version of *c* is process-connected to *e*. To see this, let **S** (for sufficient) be any set of events satisfying the following two conditions:

i) all the events in **S** occur in @ at *t*, and

ii) **S** is a sufficient set for *e*.

We first prove that *c* or a more or less fragile version of *c* belongs to **S**. To see this, let *w* be an arbitrary world among the closest-at-*t* not-*c*-worlds in $\mathscr{H}$. Suppose for *reductio* that all the events in **S** occur in *w*. Since **S** is a sufficient set for *e*, it follows that *e* occurs in *w*. But since *e* depends counterfactually on *c* within $\mathscr{H}$, we already know that *e* does not occur in *w*. Thus, the supposition leads to a contradiction. And so we conclude that at least one event in **S** does not occur in *w*. Since *c* and more or less fragile versions of *c* are the only events that occur in @ at *t*, but do not occur in *w*, it follows that *c*, or a more or less fragile version of *c*, belongs to **S**.

The above argument applies to any set **S** of events satisfying the two conditions i) and ii) above. Now, let $\mathscr{T}$ (for total) be the set of all those events that occur in @ at *t* and are process-connected to *e*. To prove that *c*, or a more or less fragile version of *c*, is process-connected to *e*, it is enough to prove that $\mathscr{T}$ satisfies conditions i) and ii).

To see this, we note that our definition of $\mathscr{T}$ – as the set of all those events that occur in @ at *t* and are process-connected to *e* – ensures that condition i) is satisfied. And, given the *Assumption of sufficiency*, condition ii) is also satisfied. Hence *c*, or a more or less fragile version of *c*, is process-connected to *e*.

1.2 The assumption of sufficiency

The above proof rests on the *Assumption of sufficiency*, stated at the beginning of section 1.1. For ease of reference, I repeat the assumption here:

> *Assumption of sufficiency*: let e be an instantaneous event that occurs in @ at t', let t be some time strictly earlier than t', and let $\mathscr{T}$ be the set of all those events that occur in @ at t and are process-connected to e. Then $\mathscr{T}$ is sufficient for e.

I cannot offer a general proof of this assumption. In the following, however, I will show that the assumption is true for all neuron diagrams, given our standard restriction on the domain of neuron events to include only maximally temporally fragile events (cf. Chapter 4 section 2.3). This makes it plausible that the assumption is satisfied in most or all ordinary cases.

To see that the *Assumption of sufficiency* holds in the case of neuron diagrams, let the time-series T: $t_1 < t_2 < \ldots < t_n$ be a time-series that includes all those times at which it is explicitly represented what the neurons are doing. Let e_n be the effect we are interested in, and suppose that e_n occurs at time t_n. Let $\mathscr{T}_1$ be the set containing all those events that occur at time t_1 and are process-connected to e_n. Then we must prove that $\mathscr{T}_1$ is sufficient for e_n. To do this, we shall work our way back stepwise from t_n to t_1.

For the first step, let $\mathbf{S}_{n-1}$ be the set of *all* the neuron events occurring at time t_{n-1}. Together, these events fully characterise the complete state of our neuron world at time t_{n-1}. Given forwards determinism, it follows that $\mathbf{S}_{n-1}$ is sufficient for e_n. From the definition of minimal sufficiency, it now follows that there exists at least one set S_{n-1} of contemporaneous events that occurs at t_{n-1} and is minimally sufficient for e_n (where $\mathbf{S}_{n-1}$ is either identical to S_{n-1}, or a strictly more fragile version of S_{n-1}). And since e_n is already maximally temporally fragile, it immediately follows that S_{n-1} is time-sensitively sufficient for e_n. From this, it follows that for every event e_{n-1} in S_{n-1}, there is an apparent process from e_{n-1} to e_n. And since the time-series T includes all those times at which it is explicitly represented what the neurons are doing, it immediately follows that there is a genuine process from e_{n-1} to e_n. Thus, e_{n-1} is process-connected to e_n.

Let $\mathcal{T}_{n-1}$ be the set of all those neuron events that occur at time t_{n-1} and are process-connected to e_n. It follows that S_{n-1} is a (proper or improper) subset of $\mathcal{T}_{n-1}$. And since S_{n-1} is sufficient for e_n, it follows that $\mathcal{T}_{n-1}$ is sufficient for e_n.

We may now apply exactly the same argument to every event e_{n-1} that belongs to $\mathcal{T}_{n-1}$. Let S_{n-2} be the set of all those events that belong to a time-sensitively sufficient set for some event in S_{n-1}. By the same arguments as above, it follows that each event in S_{n-2} is process-connected to some event in S_{n-1}. By the transitivity of process-connection, which I prove in section 2 below, it follows that each event in S_{n-2} is process-connected to e_n.

Let $\mathcal{T}_{n-2}$ be the set of all those neuron events that occur at time t_{n-2} and are process-connected to e_n. It follows that S_{n-2} is a (proper or improper) subset of $\mathcal{T}_{n-2}$. By the same arguments as above, we find that $\mathcal{T}_{n-2}$ is sufficient for each event in $\mathcal{T}_{n-1}$. From this, it follows that $\mathcal{T}_{n-2}$ is sufficient for e_n. To see this, simply note that sufficiency is transitive: if every nomologically possible world where all the events in $\mathcal{T}_{n-2}$ occur is such that all the events in $\mathcal{T}_{n-1}$ occur, and every nomologically possible world where all the events in $\mathcal{T}_{n-1}$ occur is such that e_n occurs, then every nomologically possible world where all the events in $\mathcal{T}_{n-2}$ occur is such that e_n occurs. Thus, $\mathcal{T}_{n-2}$ is sufficient for e_n.

By iterating the above argument, we now find that $\mathcal{T}_1$ is sufficient for e_n. This shows that, given our usual restriction on the domain of neuron events, the *Assumption of sufficiency* is true in the case of neuron diagrams.

Since there is in principle no limitation on the number of neurons and the number of time steps that a neuron diagram may involve, neuron diagrams may accurately represent a wide range of real-life cases. This makes it highly plausible – though I cannot offer a direct proof – that the *Assumption of* *sufficiency* is satisfied in all ordinary cases, given our ordinary restrictions on the domain of instantaneous events. By extension, this means that we can apply the principle of *Restricted sufficiency of counterfactual dependence for process-connection* in such ordinary cases.

2. The transitivity of process-connection

In this section, I show that the definition of process-connection implies that process-connection is a transitive relation:

> *Transitivity of process-connection*: if there is a set of events $\{d_1, d_2, \ldots, d_n\}$, such that c is process-connected to d_1, d_1 is process-connected to $d_2, \ldots$, and d_n is process-connected to e, then c is process-connected to e.

To show this, I show that for any three events *c*, *d*, and *e*, it is the case that if *c* is process-connected to *d*, and *d* is process-connected to *e*, then *c* is process-connected to *e*. The above statement of the transitivity of process-connection follows immediately from this result.

Let *c*, *d*, and *e* be three events, such that *c* is process-connected to *d*, and *d* is process-connected to *e*. This implies that there is a series of events $P(c, d)$ characterising a genuine process from *c* to *d*, and a series of events $P(d, e)$ characterising a genuine process from *d* to *e*. Let $T(c, d)$ be the time-series on which $P(c, d)$ is based, and let $T(d, e)$ be the time-series on which $P(d, e)$ is based. Finally, let $P(c, e)$ be the series of events that results from letting $P(c, d)$ continue into $P(d, e)$, removing the repetition of *d*. And let $T(c, e)$ be the time-series that results from letting $T(c, d)$ continue into $T(d, e)$, removing the repetition of *d*'s time of occurrence.

I will now show that $P(c, e)$ characterises a genuine process from *c* to *e*. To see this, note first that it follows from the fact that $P(c, d)$ is based on $T(c, d)$, and $P(d, e)$ is based on $T(d, e)$, that $P(c, e)$ is based on $T(c, e)$. Furthermore, it is clear that $P(c, e)$ characterises an apparent process: each nonfinal event in $P(c, e)$ belongs to a set of events that is time-sensitively sufficient for the next event in $P(c, e)$. The remaining question is: does $P(c, e)$ also characterise a genuine process?

For ease of reference, I repeat the definition of a genuine process here (cf. Chapter 6 section 4):

> *Genuine process*: a series of events P based on a time-series T characterises a *genuine process* iff P belongs to a set $\mathbf{P}$ of series of events, such that
>
> a) there is a one-one mapping between $\mathbf{P}$ and the master-set $\mathbf{T}$ for T that maps a series of events P_i from $\mathbf{P}$ to a time-series T_i in $\mathbf{T}$ iff P_i is based on T_i,
> b) each series of events in $\mathbf{P}$ characterises an apparent process, and
> c) for any time t, all series of events in $\mathbf{P}$ that associate an event with t associate the same event with t.

Let $\mathbf{P}(c, d)$ be the set of series of events in virtue of which $P(c, d)$ characterises a genuine process. And let $\mathbf{P}(d, e)$ be the set of series of events in virtue of which $P(d, e)$ characterises a genuine process. It follows from the definition of a genuine process that for each augmented version $T(c, d)^+$ of $T(c, d)$, the set $\mathbf{P}(c, d)$ includes exactly one series of events based on $T(c, d)^+$. Similarly, for each augmented version $T(d, e)^+$ of $T(d, e)$, the set $\mathbf{P}(d, e)$ includes exactly one series of events based on $T(d, e)^+$.

We may now construct a set of series of events, $\mathbf{P}(c, e)$, such that $\mathbf{P}(c, e)$ includes a series of events just in case this series of events is the result of letting a series of events that belongs to $\mathbf{P}(c, d)$ continue into a series of events that belongs to $\mathbf{P}(d, e)$, removing the repetition of d. I will now show that $P(c, e)$ characterises a genuine process in virtue of the set $\mathbf{P}(c, e)$:

First, it is obvious that $P(c, e)$ itself belongs to $\mathbf{P}(c, e)$, since $P(c, e)$ is the result of letting $P(c, d)$, which belongs to $\mathbf{P}(c, d)$, continue into $P(d, e)$, which belongs to $\mathbf{P}(d, e)$, removing the repetition of d. In addition, we find that the three conditions a), b), and c), are satisfied:

Condition a)

Let **T**(c, e) be the master-set for $T(c, e)$. By the definition of a master-set, a time-series $T(c, e)^+$ belongs to **T**(c, e) just in case $T(c, e)^+$ is an augmented version of $T(c, e)$. It is now crucial to note that any augmented version, $T(c, e)^+$, of $T(c, e)$ may be obtained by letting an augmented version $T(c, d)^+$ of $T(c, d)$ continue into an augmented version $T(d, e)^+$ of $T(d, e)$, removing the repetition of d's time of occurrence. For each such time-series $T(c, e)^+$, the set **P**(c, e) contains exactly one series of events based on $T(c, e)^+$ – namely, the series of events that is the result of letting the unique series of events associated with the corresponding augmented version $T(c, d)^+$ of $T(c, d)$ in **P**(c, d) continue into the unique series of events associated with the corresponding augmented version $T(d, e)^+$ of $T(d, e)$ in **P**(d, e), removing the repetition of d. Thus, condition a) is satisfied.

Condition b)

It is clear that each series of events in **P**(c, e) characterises an apparent process.

Condition c)

From the fact that $P(c, d)$ characterises a genuine process, it follows that for any time t, all series of events in **P**(c, d) that associate an event with t associate the same event with t. From the fact that $P(d, e)$ characterises a genuine process, it similarly follows that for any time t, all series of events in **P**(d, e) that associate an event with t associate the same event with t. Given our construction of **P**(c, e), it follows from this that for any time t, all series of events in **P**(c, e) that associate an event with t associate the same event with t.

This concludes my proof of the *Transitivity of process-connection*.

3. The intrinsicness of process-connection

In this section, I show that process-connection is intrinsic to a process:[1]

> *Intrinsicness of process-connection*: let c and e be events occurring in world w.
> Let c be process-connected to e, and let $\mathscr{S}$ be a structure of events occurring in w, including all the events involved in a genuine process from c to e.
> Let $\mathscr{S}^*$ be a structure of events occurring in w^*, such that
> a) w and w^* are governed by the same laws, and
> b) $\mathscr{S}^*$ is the image of $\mathscr{S}$ under a structure-preserving mapping M.
> Then the image c^* of c under M is *process-connected* to the image e^* of e under M.

For ease of reference, I repeat the definitions of what it is for an event to be the image of another, and what it is for a structure of events to be the image of another (cf. Chapter 11 section 3.1):

> *The image of an event e*: let M be a structure-preserving mapping from time-slices in a world w to time-slices in a world w^*.
> Then $e^* = (w^*, t^*, s^*, I^*, C^*)$ is the image of $e = (w, t, s, I, C)$ under M, iff
> a) t^* = the image of t under M,
> b) I^* = the image of I under M, and
> c) $C^* = C$.

> *The image of a structure of events*: let M be a structure-preserving mapping from time-slices in a world w to time-slices in a world w^*.
> Then a structure $\mathscr{S}^*$ of events occurring in w^* is the image under M of a structure $\mathscr{S}$ of events occurring in w iff:
> an event e belongs to $\mathscr{S}$ just in case the image e^* of e under M belongs to $\mathscr{S}^*$.

[1] Note that I have stated my definitions of process-connection, etc., from the standpoint of the actual world. To apply them to the worlds w and w^*, as the intrinsicness principle requires us to do, we may simply let w or w^*, as appropriate, play the role of the actual world (cf. Lewis (1986*d*), p. 163).

To establish the intrinsicness of process-connection, the first step is to set out precisely what it takes for a structure of events to include all the events involved in a genuine process from *c* to *e*, as applied to real-life cases as well as neuron diagrams (section 3.1). Next, I prove two auxiliary results (section 3.2). And finally, I give my proof of the *Intrinsicness of process-connection* (section 3.3).

3.1 Including all events involved in a genuine process

A structure of events is simply a set of events. Which events must be included in a structure of events $\mathscr{S}$ in order for $\mathscr{S}$ to include all the events that are involved in a genuine process from *c* to *e*?

As we have seen (in Chapter 11 section 3.1), we may think of the intrinsicness principle as a supervenience claim. In the case of process-connection, this amounts to the claim that the relation of process-connection between *c* and *e* supervenes on any structure of events $\mathscr{S}$ that includes *all* events involved in a genuine process from *c* to *e*. In spelling out which events must be included in such a structure of events, we of course need to respect our intuitive judgement that certain events are not at all involved in a genuine process from *c* to *e*. Within these (somewhat flexible) boundaries, however, we may now do some reverse engineering in order to arrive at a definition that makes the supervenience claim come out true.

Suppose there is a genuine process from *c* to *e*. Then it is clear that our structure of events $\mathscr{S}$ needs to include all the events that belong to some characterisation *P* of a genuine process from *c* to *e*.[2] To ensure that the supervenience claim comes out true, however, our structure of events sometimes needs to include *more* events than just the events belonging to some such characterisation *P*.

First, *P* may characterise a genuine process from *c* to *e* on the basis of a time-series that includes just a few times. However, our structure of events $\mathscr{S}$

[2] In some cases, there may be more than one genuine process from *c* to *e*. For this reason, I do not say *the* genuine process from *c* to *e*.

needs to include the relevant events at *every* time from the occurrence of *c* to the occurrence of *e*. Second, it may be the case that the events in *P combine* with other events to satisfy the definition of characterising a genuine process. In that case, these other events need to be included in $\mathcal{S}$ as well.[3] In the following, I define what it takes for a structure of events $\mathcal{S}$ to include all the events that are involved, in this wider sense, in a process from *c* to *e*.

When a series of events *P* characterises a genuine process, it does so in virtue of the fact that *P* belongs to a set of series of events **P** of a certain kind (cf. Chapter 6 section 4):

> *Genuine process*: a series of events *P* based on a time-series *T* characterises a *genuine process* iff *P* belongs to a set **P** of series of events, such that
>
> a) there is a one-one mapping between **P** and the master-set **T** for *T* that maps a series of events P_i from **P** to a time-series T_i in **T** iff P_i is based on T_i,
> b) each series of events in **P** characterises an apparent process, and
> c) for any time *t*, all series of events in **P** that associate an event with *t* associate the same event with *t*.

This yields a first requirement on our structure of events $\mathcal{S}$: given that *c* is process-connected to *e*, there is a series of events *P* characterising a genuine process from *c* to *e*. By the above definition, this means that there is a set of series of events **P** in virtue of which *P* characterises a genuine process. Our requirement now is that $\mathcal{S}$ must be such that, for some set of series of events **P** in virtue of which *P* characterises a genuine process, it is the case that any event belonging to a series of events P^+ in **P** also belongs to $\mathcal{S}$. This ensures that all the times between *c*'s time of occurrence and *e*'s time of occurrence are accounted for.

In addition, $\mathcal{S}$ needs to include the events in virtue of which each series

[3] Cf. Hall (2002), pp. 264-65; Paul and Hall (2013), p. 126.

of events P^+ in **P** characterises an apparent process. To illustrate, consider two events, e_i and e_{i+1}, such that e_i is followed immediately by e_{i+1} in a series of events P^+. From the definition of an apparent process (see Chapter 6 section 4), it follows that e_i belongs to a set of contemporaneous events S_i, such that S_i is time-sensitively sufficient for e_{i+1}. This means that S_i is minimally sufficient for e_{i+1} *and* for every temporally more fragile version ${e_{i+1}}^+$ of e_{i+1}, there is a more fragile version S_i^+ of S_i, such that S_i^+ is minimally sufficient for ${e_{i+1}}^+$. It is in virtue of S_i and these more fragile versions of S_i that e_i satisfies the requirement of belonging to a time-sensitively sufficient set for e_{i+1}.

To ensure that $\mathscr{S}$ is sufficiently comprehensive to support the supervenience claim, we therefore need to impose a second requirement. To capture this requirement, it will be useful to begin with the following definition:

> *Basis for time-sensitive sufficiency*: a set $\mathbf{S}_i$ of contemporaneous events is a basis for the fact that an event e_i belongs to a time-sensitively sufficient set for a later event e_{i+1}, iff there is a set S_i such that
>
> a) e_i belongs to S_i, and
>
> b) for every temporally more fragile version ${e_{i+1}}^+$ of e_{i+1},
> there is a more fragile version S_i^+ of S_i, such that
>
>   i) S_i^+ is minimally sufficient for ${e_{i+1}}^+$, and
>
>   ii) S_i^+ is a subset of $\mathbf{S}_i$.

In the case of neuron diagrams, any set S_i that is minimally sufficient for an event e_{i+1} is *ipso facto* time-sensitively sufficient for e_{i+1}, since the neuron events we consider are already maximally temporally fragile. In the case of neuron diagrams, we therefore find that whenever an event e_i belongs to a set S_i that is minimally sufficient for e_{i+1}, then S_i is also a basis for the fact that e_i belongs to a time-sensitively sufficient set for e_{i+1}.

In real-life cases, however, there are usually many events that matter for the precise timing of e_{i+1}, even though they do not belong to a time-sensitively sufficient set for e_{i+1}. In our standard late preemption case where Suzy and Billy both throw rocks at the window, for example, we find that {*Suzy's throw*} is time-sensitively sufficient for *Window-shattering*. However, the basis for this fact is a much larger set of events that includes both a more fragile version of Suzy's throw, and a fragile version of Billy's throw, since the air currents and gravitational effects from Billy's rock also have some minuscule effect on the timing of the window-shattering.

To ensure that a structure $\mathscr{S}$ includes all the events involved in a genuine process from c to e, we may now use the above definition to impose the following further requirement: for any two events, e_i and e_{i+1}, such that e_i is followed immediately by e_{i+1} in a series of events P^+ in $\mathbf{P}$, it must be the case that some basis $\mathbf{S}_i$ for the fact that e_i belongs to a time-sensitively sufficient set for e_{i+1} is a subset of $\mathscr{S}$.

Based on these requirements, we may now define what it takes for a structure of events $\mathscr{S}$ to include all events involved in a genuine process from c to e:

Including all events involved in a genuine process:
a structure of events $\mathscr{S}$ includes all the events in a genuine process from c to e iff there is a set of series of events $\mathbf{P}$, such that

a) some series of events P characterises a genuine process from c to e in virtue of belonging to $\mathbf{P}$,
b) if an event belongs to a series of events P^+ in $\mathbf{P}$, then it belongs to $\mathscr{S}$, and
c) for any two events, e_i and e_{i+1}, such that e_i is followed immediately by e_{i+1} in a series of events P^+ in $\mathbf{P}$, it is the case that some basis $\mathbf{S}_i$ for the fact that e_i belongs to a time-sensitively sufficient set for e_{i+1} is a subset of $\mathscr{S}$.

As we shall prove in section 3.3, any structure of events satisfying this definition is sufficiently comprehensive to make the intrinsicness of process-connection come out true.

Of course, some structures that satisfy the above definition may be much more comprehensive than others: the above definition does not include any kind of minimality requirement. In practice, we may often prefer to work with structures that are also minimal, i.e. that do not include a structure satisfying the definition as a proper subset. However, such a minimality requirement is not needed as part of the definition of what it takes for a structure of events to include all the events involved in a genuine process from c to e.

Finally, the above definition is equivalent to the following much simpler definition in the case of neuron diagrams:

> *Including all events involved in a genuine process – applied to neuron diagrams*:
> a structure of events $\mathscr{S}$ includes all the events in a genuine process from c to e iff there is a series of events P, based on a time-series that includes all those times at which the neuron diagram explicitly represents what happens, such that
>
> a) P characterises a genuine process from c to e,
> b) if an event belongs to P, then it belongs to $\mathscr{S}$, and
> c) for any two events e_i and e_{i+1}, where e_i is followed immediately by e_{i+1} in P, e_i belongs to a set S_i of contemporaneous events, such that
>   i) S_i is minimally sufficient for e_{i+1}, and
>   ii) S_i is a subset of $\mathscr{S}$.

This result is based on two features of neuron diagrams that we have already encountered: first, if a series of events P characterises an apparent process and is based on a time-series that includes all those times at which the neuron diagram explicitly represents what happens, then P characterises a genuine process (cf. Chapter 6 section 4). And second, if a set S_i of events is minimally sufficient for a later event e_{i+1}, then S_i is time-sensitively sufficient for e_{i+1}, since our neuron events are already maximally temporally fragile.

In the case of neuron diagrams, it is therefore reasonably simple to identify a structure of events that includes all the events involved in a genuine process from c to e.

3.2 Two auxiliary results

In this section, I prove two auxiliary results that will be useful in the proof of the *Intrinsicness of process-connection*. The first result is:[4]

> *Auxiliary result I*: a set S_1 of contemporaneous events is sufficient for a later event e_2 iff the image S_1* of S_1 is sufficient for the image e_2* of e_2.

For ease of reference, I repeat the definition of sufficiency here (cf. Chapter 6 section 2):

> *Sufficiency*: a set S_1 of contemporaneous events is *sufficient* for an event e_2 iff
> a) S_1 occurs strictly earlier than e_2, and
> b) for every nomologically possible world w, it is the case that if every event in S_1 occurs in w, then e_2 occurs in w.

Note that we obtain the image $e*$ of an event e under M simply by shifting the time t and interval I on which e is based, according to the structure-preserving mapping M. Importantly, the class of complete states C on which e is based is held fixed. Similarly, we obtain the image $S*$ of a set S of contemporaneous events by finding the image $e*$ of each event e in S. Thus, the image S_1* of a set of contemporaneous events S_1 is itself a set of contemporaneous events, and

[4] Here, and in the following, I shall avoid repeated relativisations to world w and $w*$, with the understanding that events (and sets of events) denoted *without* an asterisk occur in world w, while events (and sets of events) denoted *with* an asterisk occur in $w*$, and are images, under a given structure-preserving mapping M, of the corresponding events (and sets of events) in w.

the only difference between S_1 and S_1* is that the relevant time and intervals have been shifted, according to the structure-preserving mapping.

To show the first direction, suppose that S_1 is sufficient for e_2. We must now show that S_1* is sufficient for e_2*. Since S_1 is a set of contemporaneous events, it immediately follows that the image S_1* of S_1 is also a set of contemporaneous events. Since S_1 occurs strictly earlier than e_2, it follows – given that M is a structure-preserving mapping – that S_1* occurs strictly earlier than e_2*. Finally, it follows from the fact that S_1 is sufficient for e_2 that for every nomologically possible world w, it is the case that if every event in S_1 occurs in w, then e_2 occurs in w. From this together with the time-translation invariance of the laws, it follows that for every nomologically possible world w, it is the case that if every event in S_1* occurs in w, then e_2* occurs in w. Thus, we find that S_1* is sufficient for e_2*.

To show the other direction, suppose that S_1* is sufficient for e_2*. We must now show that S_1 is sufficient for e_2. From the fact that M is a structure-preserving mapping, it follows that no two time-slices from w are mapped into the same time-slice in w*. This means that we can define the inverse mapping M^- from time-slices in w* to time-slices in w, such that M^- maps the time-t* slice from w* into the time-t slice in w just in case M maps the time-t slice in w into the time-t* slice in w*. This means that S_1 is the image of S_1* under M^-, and e_2 is the image of e_2* under M^-. By the same arguments as above, it therefore follows that S_1 is sufficient for e_2.

This establishes our *Auxiliary result I*. I will now move on to our second auxiliary result:

> *Auxiliary result II*: a set S^+ of contemporaneous events is a strictly more fragile version of a set S iff the image S^{+*} of S^+ is a strictly more fragile version of the image S* of S.

For ease of reference, I repeat the definition of what it takes for one set of events to be a more fragile version of another (cf. Chapter 6 section 2):

> *More fragile version of a set*: a set S^+ of contemporaneous events is a *more fragile version* of a set S of contemporaneous events iff
> a) S^+ and S have the same realization in @, and
> b) $w(S^+) \subseteq w(S)$.

Note that, as mentioned in Chapter 6 section 2, a set S^+ of contemporaneous events is a *strictly* more fragile version of a set S of contemporaneous events just in case S^+ is a more fragile version of S, and $w(S^+) \subset w(S)$.

To show the first direction, suppose that S^+ is a strictly more fragile version of S. Applied to world w, this implies that S^+ and S have the same realization in w. And as is easily verified, two sets of contemporaneous events have the same realization in a given world just in case they occur at the same time in that world. Thus, S^+ and S occur at the same time. This means that S^{+*} and S^* occur at the same time in w^*, which means that S^{+*} and S^* have the same realization in w^*. Furthermore, we know that $w(S^+) \subseteq w(S)$. Surely, what is metaphysically possible at one time is exactly the same as what is metaphysically possible at another time. Thus, shifting the time makes no difference to the relevant subset relations. And so we find that $w(S^{+*}) \subseteq w(S^*)$. This shows that S^{+*} is a strictly more fragile version of S^*.

To show the other direction, suppose that S^{+*} is a strictly more fragile version of S^*. Under the inverse mapping M^-, we now find that S^+ is the image of S^{+*} under M^-, and S is the image of S^* under M^-. By the same arguments as above, we now find that S^+ is a strictly more fragile version of S.

This establishes our *Auxiliary result II*. Given these two auxiliary results, we are now ready to prove the *Intrinsicness of process-connection*.

3.3 The proof

In this section I prove the *Intrinsicness of process-connection*. For ease of reference, I repeat the statement here:

> *Intrinsicness of process-connection*: let *c* and *e* be events occurring in world *w*.
> Let *c* be process-connected to *e*, and let $\mathscr{S}$ be a structure of events occurring in *w*, including all the events involved in a genuine process from *c* to *e*.
> Let $\mathscr{S}$* be a structure of events occurring in *w**, such that
> a) *w* and *w** are governed by the same laws, and
> b) $\mathscr{S}$* is the image of $\mathscr{S}$ under a structure-preserving mapping *M*.
> Then the image *c** of *c* under *M* is *process-connected* to the image *e** of *e* under *M*.

Suppose that *c* and *e* occur in *w*, *c* is process-connected to *e*, and $\mathscr{S}$ is a structure of events occurring in *w*, including all the events involved in a genuine process from *c* to *e*. Suppose further that $\mathscr{S}$* is a structure of events occurring in *w**, such that *w* and *w** are governed by the same laws, and $\mathscr{S}$* is the image of $\mathscr{S}$ under a structure-preserving mapping *M*. Our task now is to show that the image *c** of *c* under *M* is process-connected to the image *e** of *e* under *M*.

From the fact that $\mathscr{S}$ includes all the events involved in a genuine process from *c* to *e*, it follows (cf. section 3.1) that there is a set of series of events **P**, such that:

a) some series of events *P* characterises a genuine process from *c* to *e* in virtue of belonging to **P**,
b) if an event belongs to a series of events P^+ in **P**, then it belongs to $\mathscr{S}$, and
c) for any two events, e_i and e_{i+1}, such that e_i is followed immediately by e_{i+1} in a series of events P^+ in **P**, it is the case that some basis $\mathbf{S}_i$ for the fact that e_i belongs to a time-sensitively sufficient set for e_{i+1} is a subset of $\mathscr{S}$.

Our first task is to show that the image P^* of P under the mapping M characterises an *apparent process*.

To see this, let e_i and e_{i+1} be any two events in P, such that e_i is immediately followed by e_{i+1}. We know that $\mathscr{S}$ contains every event in P, so $\mathscr{S}$ includes e_i and e_{i+1}. It follows from this that $\mathscr{S}^*$ contains the image e_i^* of e_i under M, and the image e_{i+1}^* of e_{i+1} under M. Furthermore, we know that $\mathscr{S}$ includes a set of contemporaneous events S_i, such that e_i belongs to S_i and S_i is time-sensitively sufficient for e_{i+1}. From this it follows that $\mathscr{S}^*$ contains the image S_i^* of S_i under M. Obviously, e_i^* belongs to S_i^*. What we need to show now is that S_i^* is time-sensitively sufficient for e_{i+1}^*.

From our *Auxiliary result I* (cf. section 3.2) together with the fact that S_i is sufficient for e_{i+1}, it follows that S_i^* is sufficient for e_{i+1}^*.

To show that S_i^* is minimally sufficient for e_{i+1}^*, we now need to show that there is no set $S_i^{*\prime}$, such that S_i^* is a strictly more fragile version of $S_i^{*\prime}$, and $S_i^{*\prime}$ is a sufficient set for e_{i+1}^*. To see this, suppose for *reductio* that there is a set $S_i^{*\prime}$, such that S_i^* is a strictly more fragile version of $S_i^{*\prime}$, and $S_i^{*\prime}$ is a sufficient set for e_{i+1}^*. Let S_i' be the set of events in w, such that $S_i^{*\prime}$ is the image of S_i'. It now follows from our *Auxiliary result II* that S_i is a strictly more fragile version of S_i', and it follows from our *Auxiliary result I* that S_i' is sufficient for e_{i+1}. Thus, we find that S_i is not minimally sufficient for e_{i+1}. But we already know that S_i is minimally sufficient for e_{i+1}. Thus, the supposition leads to a contradiction, and we can therefore conclude that S_i^* is minimally sufficient for e_{i+1}^*.

To show that S_i^* is *time-sensitively* sufficient for e_{i+1}^*, note that for any two events, e_i and e_{i+1}, such that e_i is followed immediately by e_{i+1} in P, there is some basis $\mathbf{S}_i$ for the fact that S_i is time-sensitively sufficient for e_{i+1}, such that $\mathbf{S}_i$ is a subset of $\mathscr{S}$. It follows from this that $\mathscr{S}^*$ includes the image $\mathbf{S}_i^*$ of $\mathbf{S}_i$ under M. For each more temporally fine-grained version of e_{i+1}^*, we can therefore apply exactly the same arguments as above. Thus, we find that S_i^* is time-sensitively sufficient for e_{i+1}^*.

The above argument applies to any two events, e_i* and e_{i+1}*, in P*. Hence, we find that P* characterises an apparent process.

Our final step is to show that P* characterises a *genuine process*. To see this, note that it follows from the definition of $\mathscr{S}$ that for any series of events P^+ in **P**, $\mathscr{S}$ includes every event that belongs to P^+. Correspondingly, $\mathscr{S}$* includes every event that belongs to the image P^+* of P^+. Let **P*** be the set of series of events such that **P*** includes a series of events P^+* just in case P^+* is the image of a series of events P^+ belonging to **P**. Our task now is to show that P* satisfies the conditions for characterising a genuine process based on **P***. For ease of reference, I repeat the definition here (cf. Chapter 6 section 4):

> *Genuine process*: a series of events P, based on a time-series T, characterises a genuine process iff P belongs to a set **P** of series of events, such that
>
> a) there is a one-one mapping between **P** and the master-set **T** for T that maps a series of events P_i from **P** to a time-series T_i in **T** iff P_i is based on T_i,
> b) each series of events in **P** characterises an apparent process, and
> c) for any time t, all series of events in **P** that associate an event with t associate the same event with t.

It is clear that P* belongs to **P***, since P* is the image of the series of events P, and P belongs to **P**. I will now go on to show that each of the three conditions is satisfied:

Condition a) of Genuine process

Let T be the time-series on which P is based, and let **T** be the master-set for T. Let T* be the image of T under M, and let **T*** be the image of **T** under M. It follows from the fact that M is a structure-preserving mapping that **T*** is the master-set for T*. Furthermore, it is easily seen that for any series of events P^+ that belongs to **P** and is based on a time-series T^+, the image P^+* of P^+ belongs to $\mathbf{P}^*$ and is based on the image T^{+*} of the time-series T^+. From this it follows that there is a one-one mapping between $\mathbf{P}^*$ and $\mathbf{T}^*$ that maps a series of events P_i^* from $\mathbf{P}^*$ to a time-series T_i^* in $\mathbf{T}^*$ iff P_i^* is based on T_i^*.

Condition b) of Genuine process

By the very same arguments as we went through in the case of P^*, we find that for every series of events P^+ in $\mathbf{P}$, the image P^{+*} of P^+ under M characterises an apparent process.

Condition c) of Genuine process

From the fact that P characterises a genuine process based on $\mathbf{P}$, it follows that for every time t, all series of events in $\mathbf{P}$ that associate an event with t associate the same event with t. From this it follows that for every time t^*, all series of events in $\mathbf{P}^*$ that associate an event with t^* associate the same event with t^*.

Hence, P^* characterises a genuine process.

From this result, the *Intrinsicness of process-connection* immediately follows: P^* is the image of P under M. P characterises a genuine process from c to e. This means that P starts with c and ends with e. Thus, P^* starts with the image c^* of c under M and ends with the image e^* of e under M. Furthermore, we have just seen that P^* characterises a genuine process. Thus, we find that c^* is process-connected to e^*.

4. The sufficiency of counterfactual dependence for security-dependence

My aim in this section is to show that counterfactual dependence within a possibility horizon $\mathscr{H}$ implies security-dependence within that same possibility horizon $\mathscr{H}$. This result is very useful: first, it provides a deeper understanding of the relation of security-dependence as a weakened version of counterfactual dependence. Second, it makes it easy to reason about security-dependence whenever we encounter a case where one event depends counterfactually on another within the relevant possibility horizon.

The principle that counterfactual dependence implies security-dependence (within a given possibility horizon) may be stated as follows:

> *Sufficiency of counterfactual dependence for security-dependence*:
> if e depends counterfactually on c within a possibility horizon $\mathscr{H}$,
> then e security-depends on c within $\mathscr{H}$.

To see this, suppose that e depends counterfactually on c within a possibility horizon $\mathscr{H}$, and suppose further that c occurs at time t. We may now go through the following three steps to show that this entails that e security-depends on c within $\mathscr{H}$. (Note that all of the following is relativised to the chosen possibility horizon $\mathscr{H}$; to avoid clutter, I leave out these relativisations.)

Step 1: the first step is to assess e's security at t in @. It follows from the fact that e depends counterfactually on c that there is at least one world where e does not occur. From this it follows that e has at least one minimal dependence set M at t in @. If there is more than one minimal dependence set for e in @ at t, it immediately follows that each of these minimal dependence sets is distinct from Ø. If there is just one minimal dependence set M for e in @ at t, we can similarly prove that $M \neq \emptyset$.

To see this, suppose for *reductio* that $M = \emptyset$. By the definition of a minimal dependence set, it follows that in the closest-at-t world(s) where all the events in Ø fail to occur, e also fails to occur. But @ itself is the closest-at-t world where all the events in Ø fail to occur. Thus, it follows that e does not occur in @. However, from the fact that e depends counterfactually on c, it follows that e does occur in @. Thus, the supposition that $M = \emptyset$ leads to a contradiction, and we can therefore conclude that $M \neq \emptyset$.

This shows that there is a non-zero distance-at-*t* from @ to the closest-at-*t* world(s) where all the events in *M* fail to occur. Thus, *e*'s security at *t* in @ is not minimal.

Step 2: the second step is to assess *e*'s security at *t* in the closest-at-*t* world(s) where *c* does not occur. Let *w* be an arbitrarily chosen world, such that *w* is among the closest-at-*t* world(s) where *c* does not occur. From the fact that *e* depends counterfactually on *c*, it follows that *e* does not occur in *w*. We therefore find that *e* has exactly one minimal dependence set at *t* in *w* – namely, the empty set Ø. From the standpoint of *w*, the closest-at-*t* world in which all the events in Ø fail to occur is *w* itself. This shows that *e*'s security at *t* in *w* is minimal. Since *w* was arbitrarily chosen among the closest-at-*t* not-*c*-world(s), the same argument applies to all the closest-at-*t* not-*c*-world(s).

Step 3: the third and final step is to compare *e*'s security at *t* in @ with *e*'s security at *t* in the closest-at-*t* world(s) where *c* does not occur. From the above results it now easily follows that *e* is *less secure* at *t* in the closest-at-*t* world(s) where *c* does not occur than it is in @.

Thus, we find, as desired, that *e* security-depends on *c* within $\mathscr{H}$. This result shows that counterfactual dependence within a possibility horizon entails security-dependence within that same possibility horizon.

B

Overview over cases

In this Appendix, I give an overview over all the cases discussed in this dissertation, together with a few further cases that I have not had space to discuss in the body of the text. In my treatment of each case, I first show the relevant neuron diagram, together with any notes about its interpretation. I then apply the two conditions of process-connection and security-dependence within each of the contextually relevant possibility horizons. I summarise the results in a table of the form:

Question	Possibility horizon	Process-connection	Security-dependence
Does A cause E?	$\mathscr{H}$	✓	✓
Does B cause E?	$\mathscr{H}$	–	✓

In the table, '✓' indicates that the condition is satisfied, and '–' indicates that it is not. On my proposed account, the two conditions of process-connection and security-dependence are individually necessary and jointly sufficient for causation – and thus, the verdict of my account can easily be read off from the table.

Note that process-connection is a binary relation, and so it is entirely independent of the chosen possibility horizon whether an instantaneous event c is process-connected to a later event e. By contrast, the relation of security-dependence is a ternary relation, and we may therefore find that e security-depends on c within a possibility horizon $\mathscr{H}_1$, but does not security-depend on c within a different possibility horizon $\mathscr{H}_2$.

The following is a complete list of the cases included in this appendix. Cases that are discussed in the body of the text are marked with an asterisk, and I have indicated in the right-hand column where they receive their main treatment.

Correlation without causation

1.*	Correlation without causation	(Chapter 7 section 1)

Redundant causation

2.*	Early preemption	(Chapter 7 section 2.1)
3.	A variant of early preemption	
4.*	Late preemption	(Chapter 7 section 2.2)
5.	Lewis-style late preemption	
6.*	A variant of trumping preemption	(Chapter 7 section 2.3)
7.*	Symmetric overdetermination	(Chapter 7 section 2.4)

Omission-involving causation

8.*	Causation by omission	(Chapter 7 section 3.1)
9.*	Prevention	(Chapter 7 section 3.2)
10.*	Double prevention	(Chapter 7 section 3.3)
11.*	Redundant causation by omission	(Chapter 7 section 3.4)
12.*	Redundant prevention	(Chapter 7 section 3.5)
13.*	Redundant double prevention	(Chapter 7 section 3.6)
14.*	Hall's case	(Chapter 7 section 3.7)

Cases that show the need for commensuration between cause and effect

15.*	Scarlet	(Chapter 7 section 4.1)
16.*	Red	(Chapter 7 section 4.2)
17.*	Apparent counterexample to transitivity	(Chapter 7 section 4.3)

Counterexamples to the transitivity of causation

18.*	Switch	(Chapter 9 section 2.2)
19.*	Boulder	(Chapter 9 section 2.3)

20.*	Switch – a modified version	(Chapter 9 section 2.4)
21.*	Boulder – a modified version	(Chapter 9 section 2.4)

Borderline cases between transitivity failure and preemption

22.*	A borderline case	(Chapter 9 section 2.5)
23.	A complex version of early preemption	
24.	Tampering	

A problem case

25.*	Case combining switch and preemption	(Chapter 9 section 2.6)

Counterexamples to the intrinsicness of causation

26.*	Switch – with right-hand track blocked	(Chapter 9 section 3.2)
27.*	Switch – with right-hand track clear	(Chapter 9 section 3.2)
28.*	Double prevention – without threat	(Chapter 9 section 3.3)

The problem of profligate omissions

29.*	The queen and the flowers	(Chapter 10 section 1)

Structurally isomorphic but causally different cases

30.*	McDermott's case	(Chapter 10 section 2)
31.	First of Hall's two structurally isomorphic cases	
32.	Second of Hall's two structurally isomorphic cases	

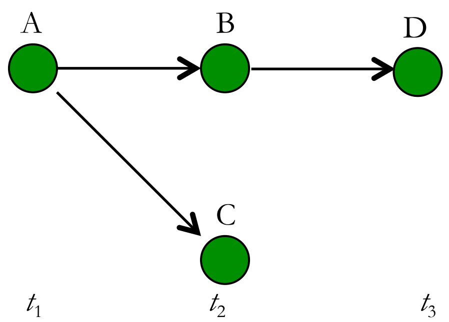
Contrastive causation

33.	Contrastive causation	(Chapter 10 section 3)

Joint causation

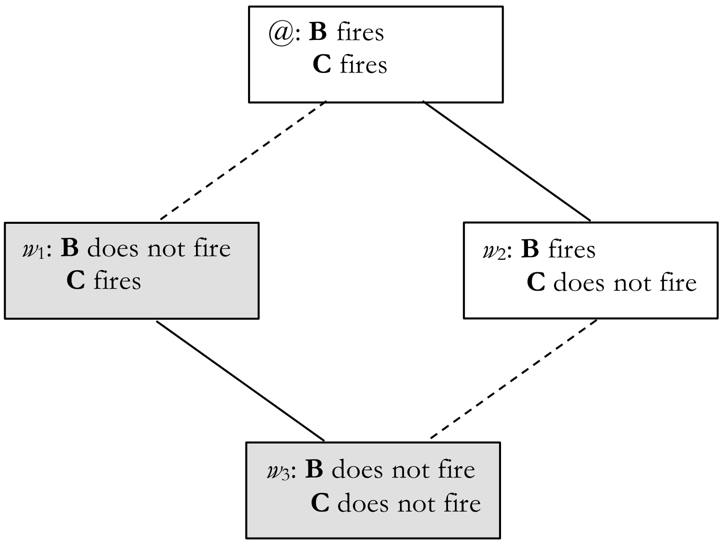
34.	Overdetermined joint causation	(Chapter 11 section 3.1)
35.	Joint causation	
36.	Complex joint causation	

1.* Correlation without causation[1]



We here consider the effect *D* and its causes at time t_2 within the possibility horizon illustrated below, where the worlds in which *D* does not happen are coloured grey:

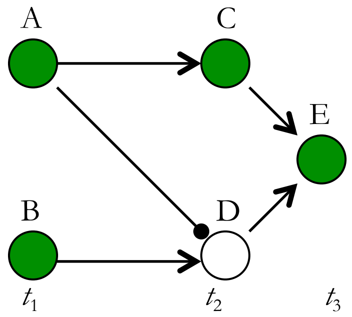
Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *B* cause *D*?	$\mathscr{H}$	✓	✓
Does *C* cause *D*?	$\mathscr{H}$	–	–

[1] Cf. Paul and Hall (2013), Figure 8.

2.* Early preemption[2]



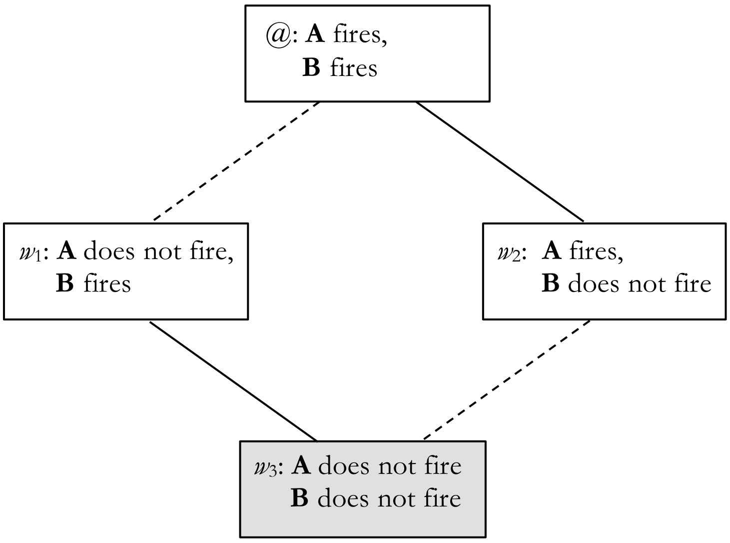
A real-life case with this structure:

> *Early preemption*: Suzy throws a rock at a window and the window shatters. If Suzy had not thrown, Billy – who is standing right next to her – would have thrown his rock, and the window would still have shattered.

We here consider the effect E and its causes at time t_1 within the possibility horizon illustrated below, where the worlds in which E does not happen are coloured grey:

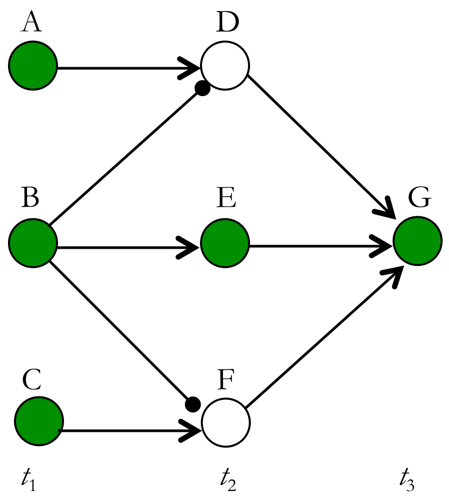
[2] Cf. Paul and Hall (2013), Figure 1.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *E*?	$\mathscr{H}$	✓	✓
Does *B* cause *E*?	$\mathscr{H}$	–	✓

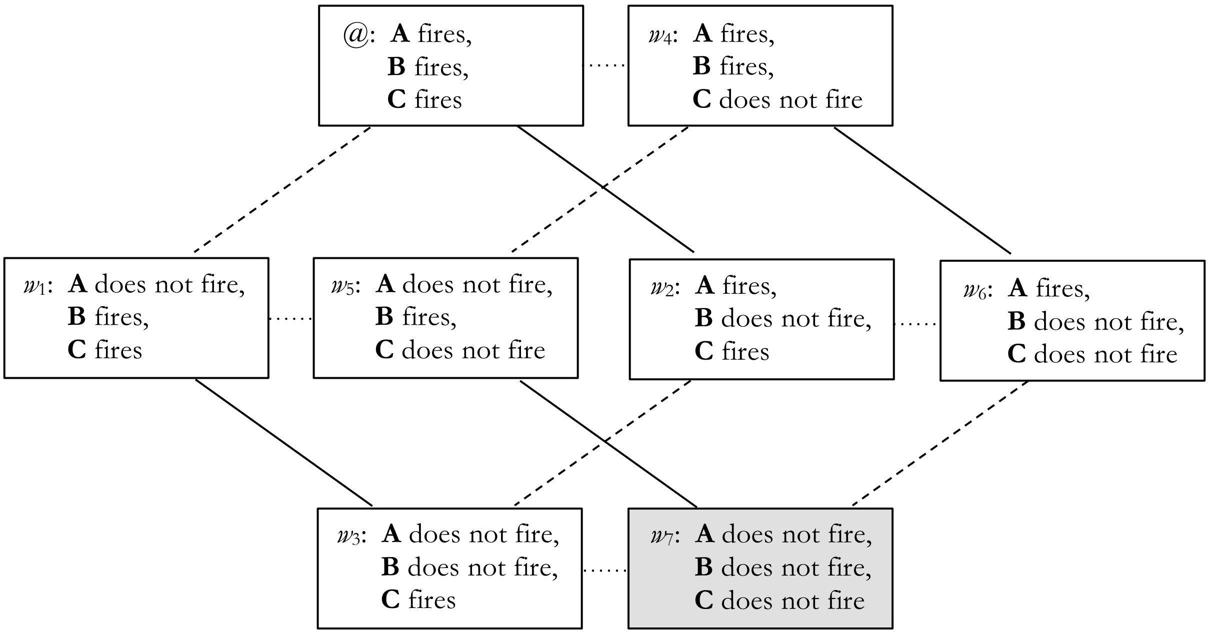
3. A variant of early preemption[3]



We here consider the effect *G* and its causes at time t_1 within the possibility horizon illustrated below, where the worlds in which *G* does not happen are coloured grey:

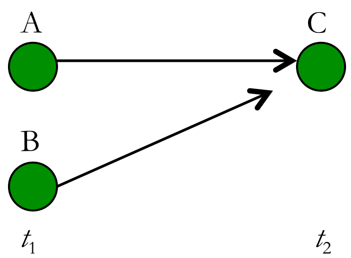
[3] Paul and Hall (2013), Figure 12.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *G*?	$\mathscr{H}$	–	✓
Does *B* cause *G*?	$\mathscr{H}$	✓	✓
Does *C* cause *G*?	$\mathscr{H}$	–	✓

4.* Late preemption[4]



Note: the speed of the stimulatory signal from **A** to **C** is independent of whether or not **B** fires. The stimulatory signal from **A** reaches **C** exactly at time t_2, whereupon **C** immediately fires, while the stimulatory signal from **B** reaches **C** at some slightly later time – say, at time t_3. The effect we are interested in – against our usual practices when dealing with neuron diagrams – is a temporally robust event, namely the event C based on the interval $I = [t_2, t_3]$, and the class of states such that **C** fires.

A real-life case with this structure:

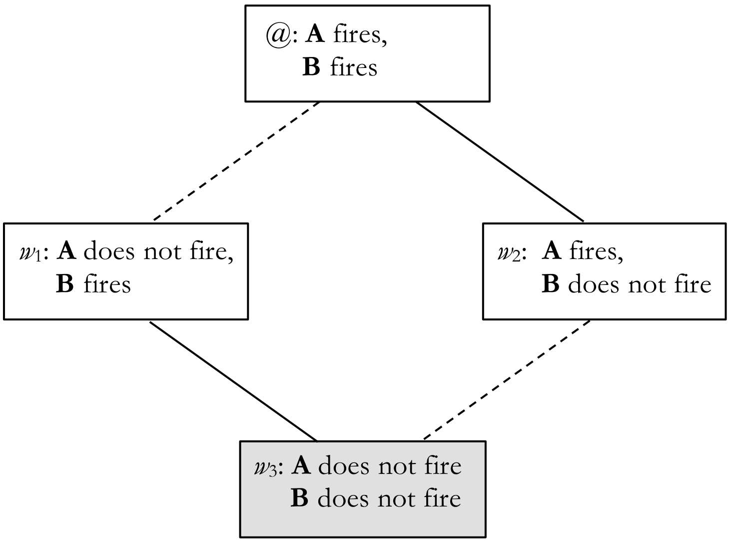
> *Late preemption*: Suzy and Billy both throw rocks at a window. Suzy's rock hits the window a moment before Billy's and the window shatters.

A here corresponds to Suzy's throw, and B corresponds to Billy's throw.

The relevant possibility horizon is as illustrated below – where the worlds in which C does not happen are coloured grey:

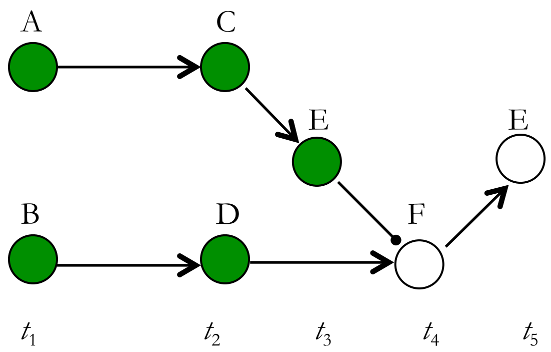
[4] Paul and Hall (2013), Figure 18.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *C*?	$\mathscr{H}$	✓	✓
Does *B* cause *C*?	$\mathscr{H}$	–	✓

5. Lewis-style late preemption[5]

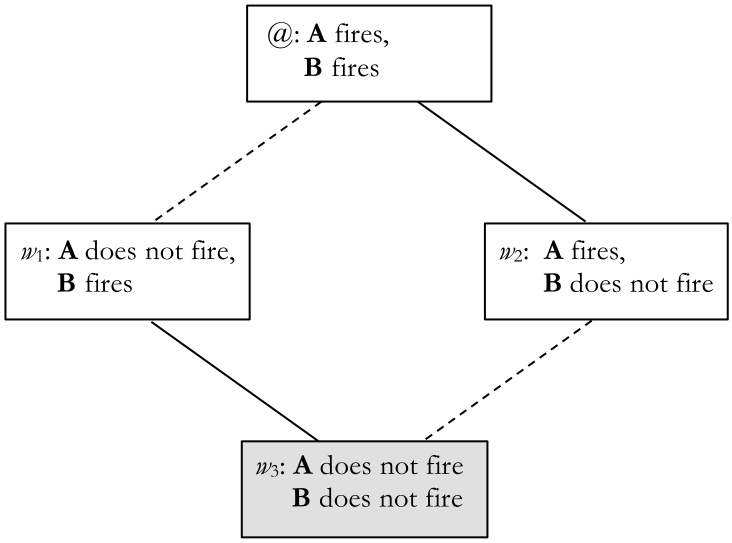


Note: Neuron **E** is represented twice – at time t_3 when it fires, and at time t_5 when it does not fire, but would have fired if *A* had not occurred. The effect we are interested in – against our usual practices when dealing with neuron diagrams – is a temporally robust event, namely the event *E* based on the interval $I = [t_3, t_5]$ and the class of states such that **E** fires.

The relevant possibility horizon is as illustrated below – where the worlds in which *E* does not happen are coloured grey:

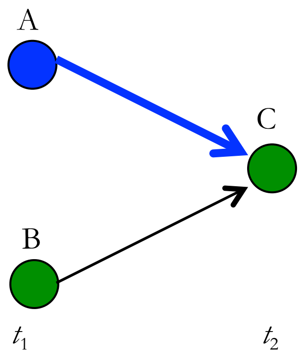
[5] This figure is closely based on Paul and Hall (2013), Figure 19. Cf. Lewis (1986*e*), p. 204.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause E?	$\mathscr{H}$	✓	✓
Does B cause E?	$\mathscr{H}$	–	✓

6.* A variant of trumping preemption[6]



Note: neuron **A** can fire in different shades, and in different intensities: it can fire in blue or green, and it can fire with intensity *I* or some other intensity. Neuron **B**, by contrast, either fires or does not fire – with no distinctions between different shades or intensities. The nomological relationships are summarised below:

	B fires	**B** doesn't fire
A doesn't fire	**C** fires	**C** doesn't fire
A fires in blue	**C** fires	**C** fires
A fires with intensity *I*, not in blue	**C** doesn't fire	**C** doesn't fire
A fires without intensity *I*, not in blue	**C** fires	**C** doesn't fire

In the case illustrated here, **A** and **B** both fire, and **A** fires in blue and with intensity *I*. We are interested in the effect *C*. At time t_1, there are five candidate causes of *C*, namely:

> *A-precise*: essentially **A**'s firing in blue and with intensity *I*
> *A-blue*: essentially **A**'s firing in blue
> *A-intensity-I*: essentially **A**'s firing with intensity *I*
> *A*: essentially **A**'s firing
> *B*: essentially **B**'s firing

[6] Cf. Paul and Hall (2013), Figure 26.

We may consider the case within two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$:
The worlds included in possibility horizon $\mathscr{H}_1$ are:

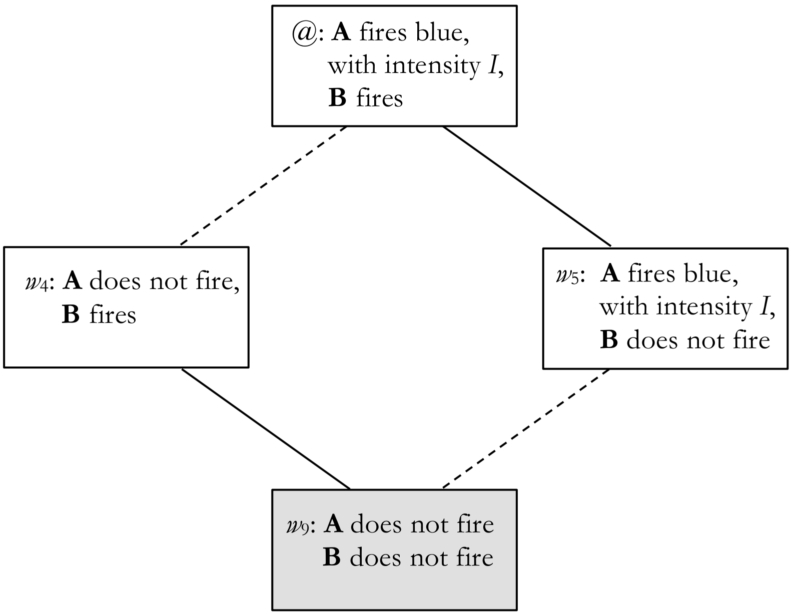
Complete state at time t_1:		*Events at t_1:*		*t_2:*
@:	**A** fires in blue, with intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_1:	**A** fires not in blue, with intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_2:	**A** fires in blue, without intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_3:	**A** fires not in blue, without intensity *I*; **B** fires	*A-blue*,	*B*	*C*
w_4:	**A** does not fire; **B** fires	*A-blue*,	*B*	*C*
w_5:	**A** fires in blue, with intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_6:	**A** fires not in blue, with intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_7:	**A** fires in blue, without intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_8:	**A** fires not in blue, without intensity *I*; **B** does not fire	*A-blue*,	*B*	*C*
w_9:	**A** does not fire; **B** does not fire	*A-blue*,	*B*	*C*

Within $\mathscr{H}_1$, we get the following results:

Question	Possibility horizon	Process-connection	Security-dependence
Does *A-precise* cause *C*?	$\mathscr{H}_1$	–	?
Does *A-blue* cause *C*?	$\mathscr{H}_1$	✓	✓
Does *A-intensity-I* cause *C*?	$\mathscr{H}_1$	–	–
Does *A* cause *C*?	$\mathscr{H}_1$	–	?
Does *B* cause *C*?	$\mathscr{H}_1$	–	–

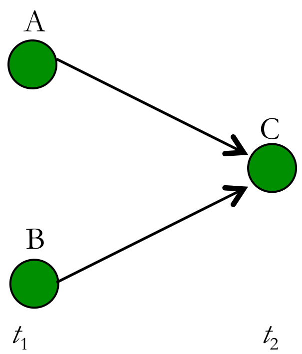
The question marks indicate that these verdicts depend on as yet unspecified details of the case.

We may also consider the case within the following more restricted possibility horizon $\mathscr{H}_2$:



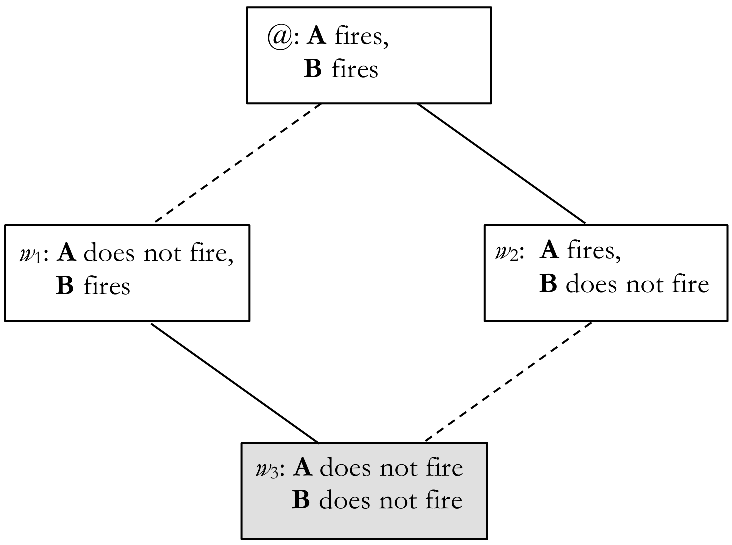
Question	Possibility horizon	Process-connection	Security-dependence
Does *A-precise* cause *C*?	$\mathscr{H}_2$	–	✓
Does *A-blue* cause *C*?	$\mathscr{H}_2$	✓	✓
Does *A-intensity-I* cause *C*?	$\mathscr{H}_2$	–	✓
Does *A* cause *C*?	$\mathscr{H}_2$	–	✓
Does *B* cause *C*?	$\mathscr{H}_2$	–	✓

7.* Symmetric overdetermination[7]



We here consider the effect *C* and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which *C* does not happen are coloured grey:

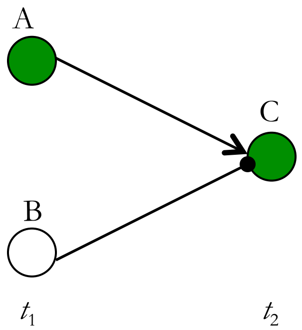
Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *C*?	$\mathscr{H}$	✓	✓
Does *B* cause *C*?	$\mathscr{H}$	✓	✓

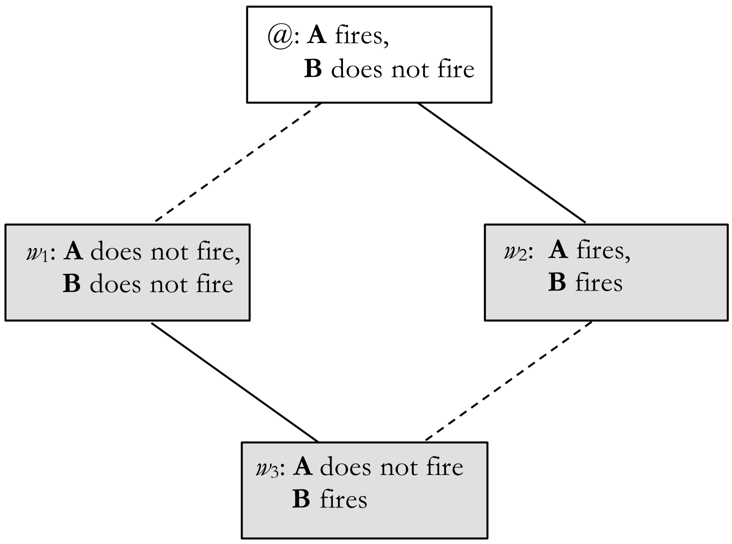
[7] Cf. Paul and Hall (2013), Figure 11.

8.* Causation by omission[8]



We here consider the effect C and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which C does not happen are coloured grey:

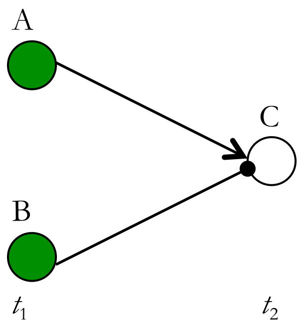
Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause C?	$\mathscr{H}$	✓	✓
Does $\neg B$ cause C?	$\mathscr{H}$	✓	✓

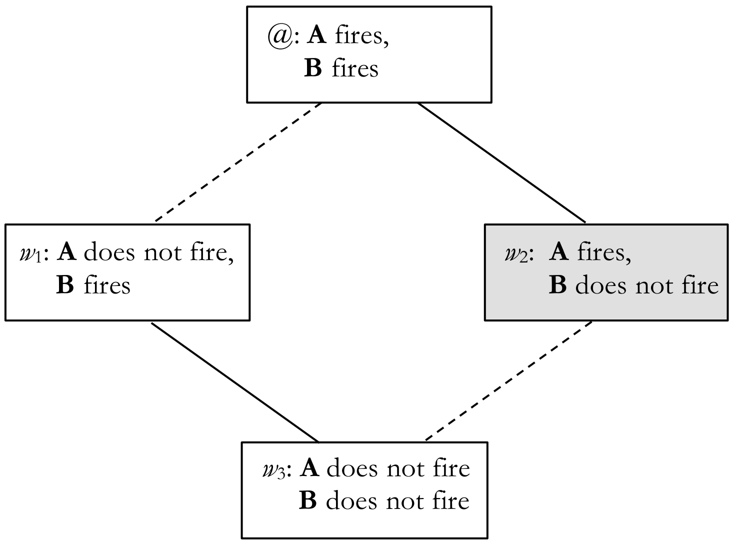
[8] Cf. Paul and Hall (2013), Figure 3.

9.* Prevention[9]



We here consider the effect $\neg C$ and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the world in which $\neg C$ does not happen is coloured grey:

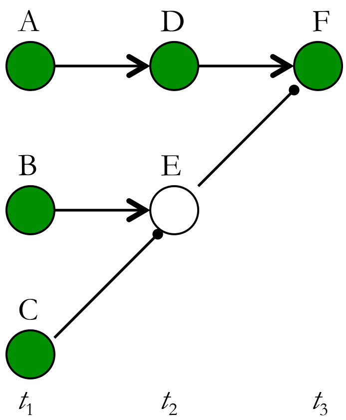
Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause $\neg C$?	$\mathscr{H}$	–	–
Does B cause $\neg C$?	$\mathscr{H}$	✓	✓

[9] Cf. Paul and Hall (2013), Figure 22.

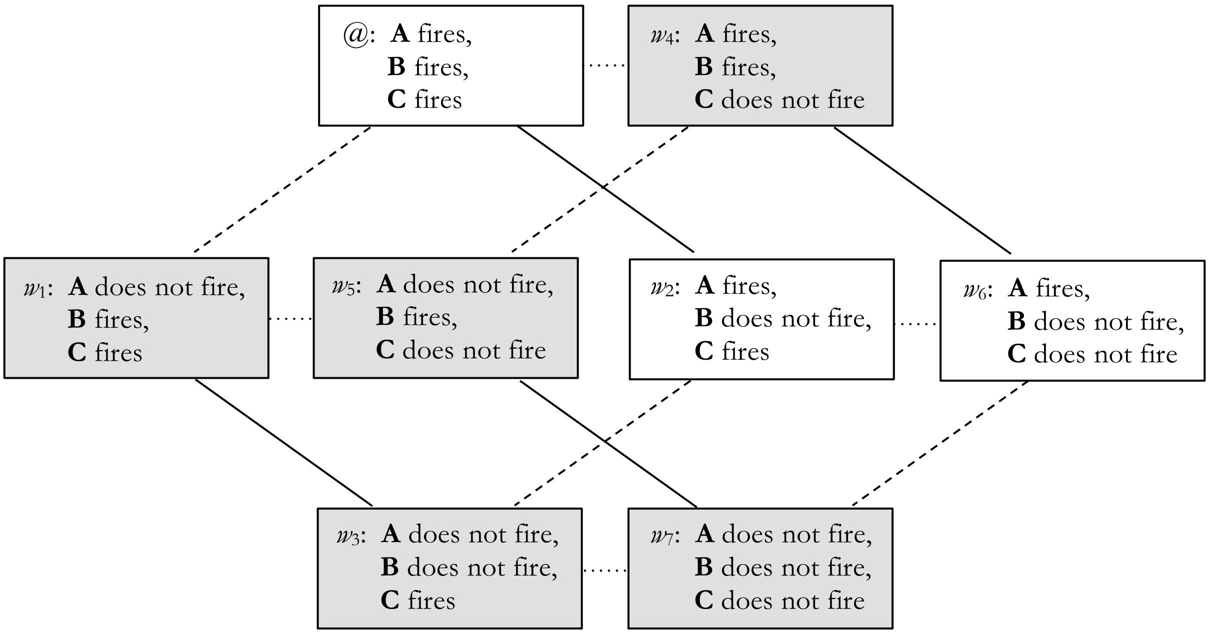
10.* Double prevention[10]



We here consider the effect *F* and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which *F* does not happen are coloured grey:

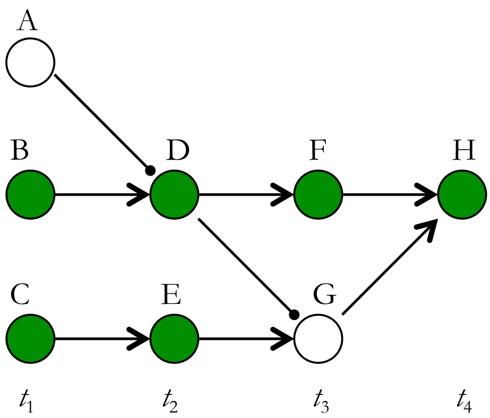
[10] Cf. Paul and Hall (2013), Figure 29.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *F*?	$\mathscr{H}$	✓	✓
Does *B* cause *F*?	$\mathscr{H}$	–	–
Does *C* cause *F*?	$\mathscr{H}$	✓	✓

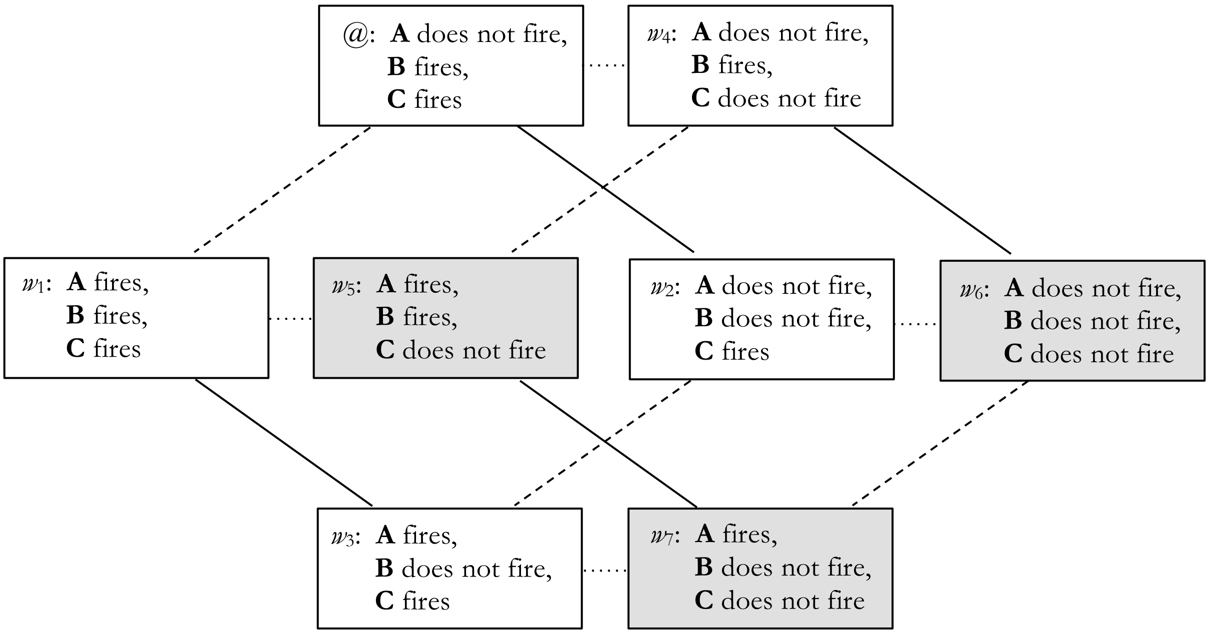
11.* Redundant causation by omission[11]



We here consider the effect *H* and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which *H* does not happen are coloured grey:

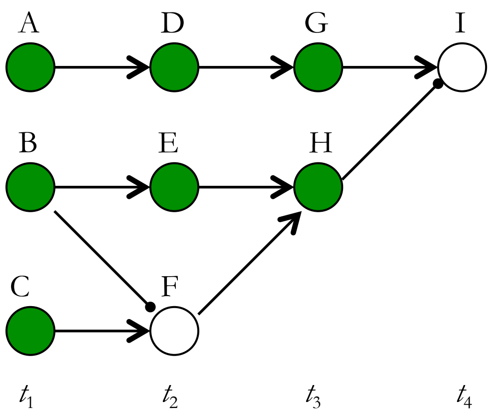
[11] Cf. Paul and Hall (2013), Figure 38.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does $\neg A$ cause H?	$\mathscr{H}$	✓	✓
Does B cause H?	$\mathscr{H}$	✓	✓
Does C cause H?	$\mathscr{H}$	–	✓

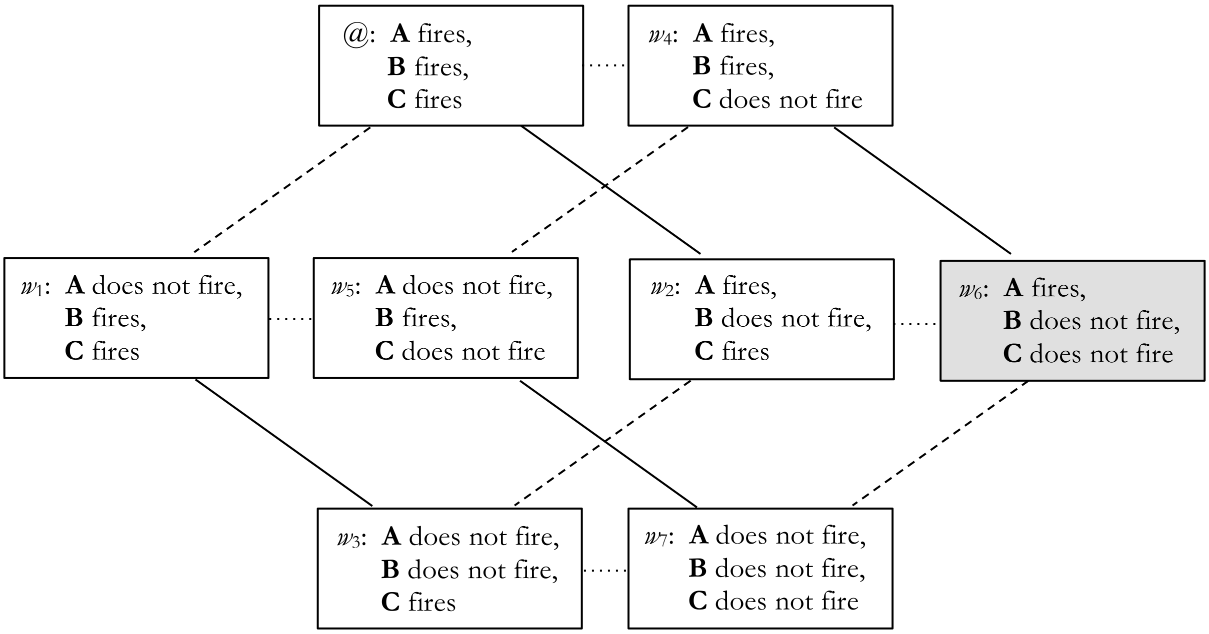
12.* Redundant prevention[12]



We here consider the effect $\neg I$ and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which $\neg I$ does not happen are coloured grey:

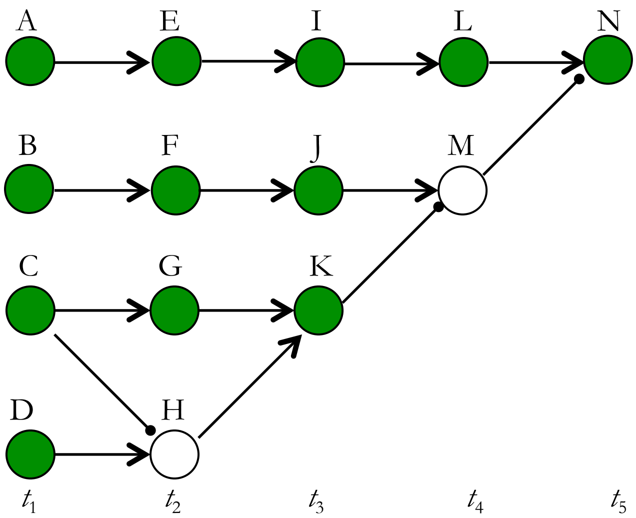
[12] Cf. Paul and Hall (2013), Figure 43.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause ¬*P*?	$\mathscr{H}$	–	–
Does *B* cause ¬*P*?	$\mathscr{H}$	✓	✓
Does *C* cause ¬*P*?	$\mathscr{H}$	–	✓

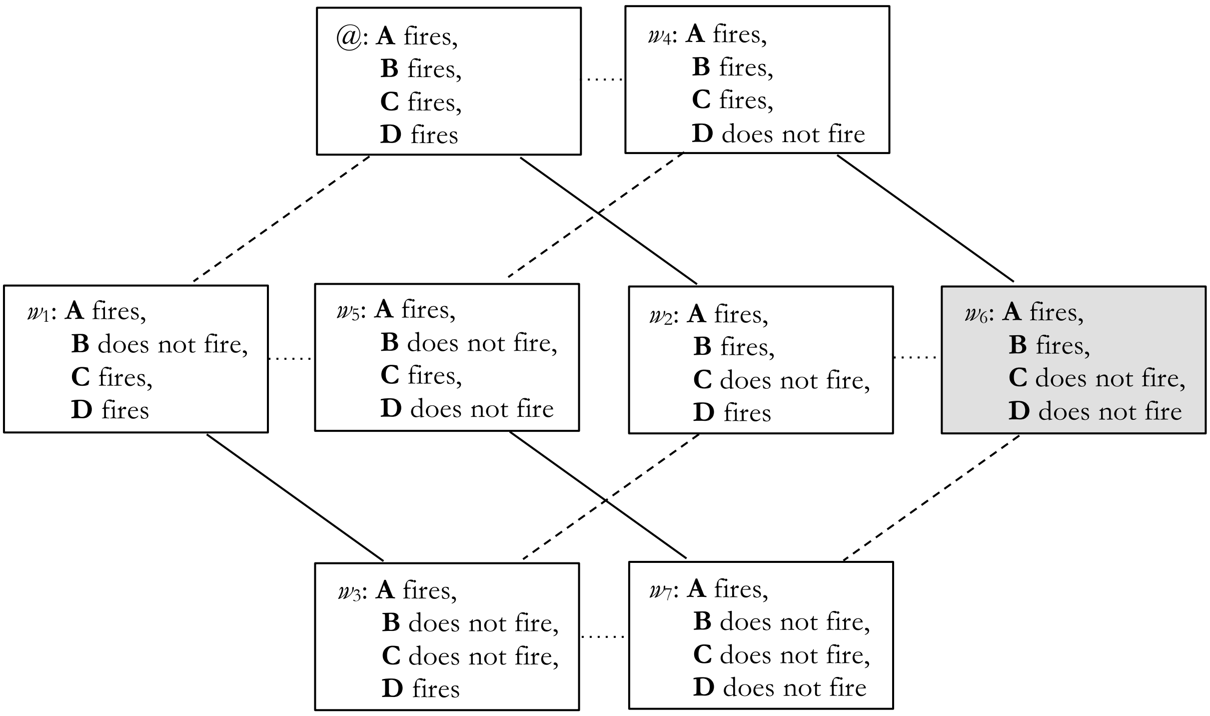
13.* Redundant double prevention[13]



We here consider the effect *N* and its causes at time t_1. For the sake of simplicity, I here treat *A* as a default event (doing so does not change the verdicts concerning *B*, *C*, and *D*). Given this, the relevant possibility horizon is as illustrated below – where the worlds in which *N* does not happen are coloured grey.

[13] Cf. Paul and Hall (2013), Figure 44.

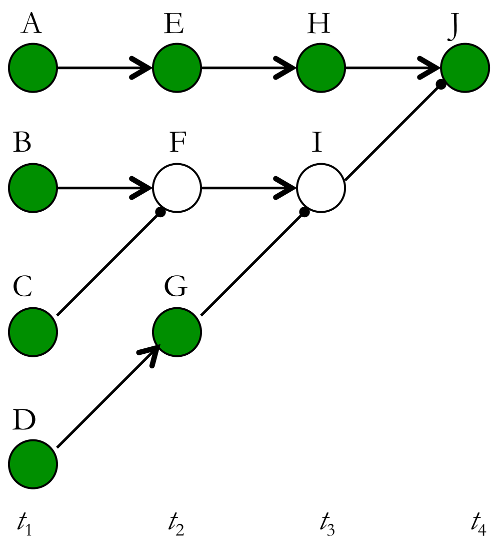
Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *B* cause *N*?	$\mathscr{H}$	–	–
Does *C* cause *N*?	$\mathscr{H}$	✓	✓
Does *D* cause *N*?	$\mathscr{H}$	–	✓

Note that, for the sake of simplicity, *A* is here treated as a background condition.

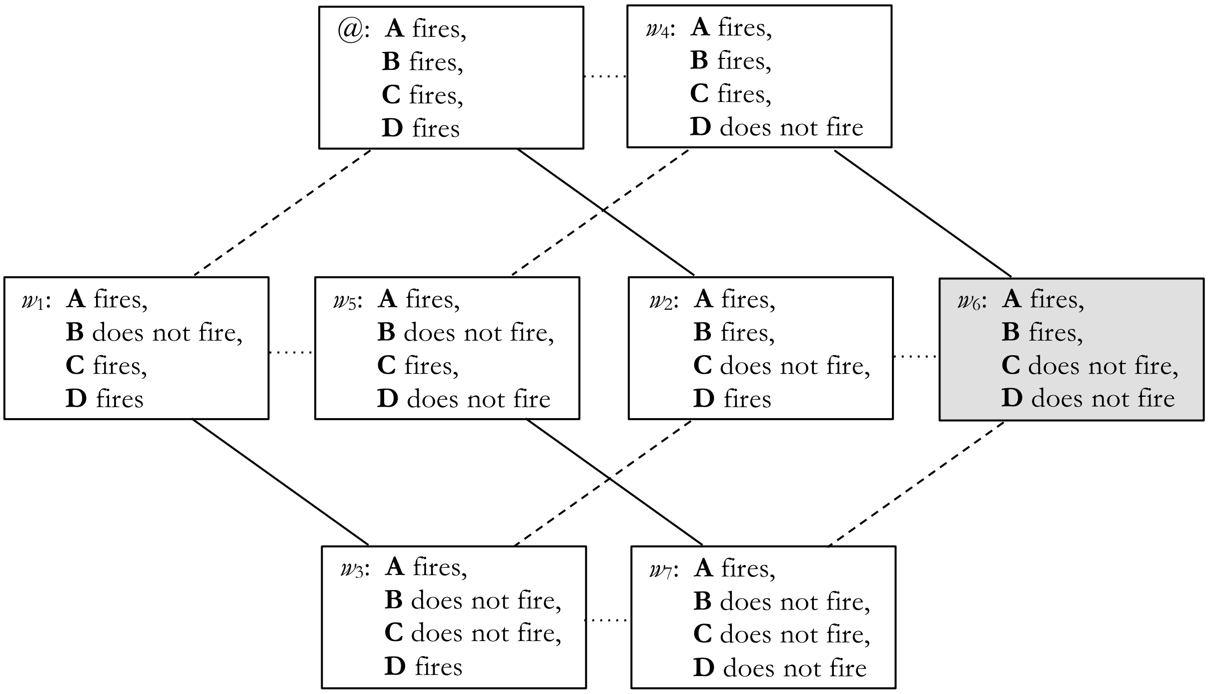
14.* Hall's case[14]



We here consider the effect *J* and its causes at time t_1. Treating *A* as a default event, we may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$ (note that treating *A* as a default event does not change the verdicts on *B*, *C*, and *D*). In my representation of both of these possibility horizons, the worlds where *J* does not happen are coloured grey:

[14] Cf. Hall (2007*b*), p. 52.

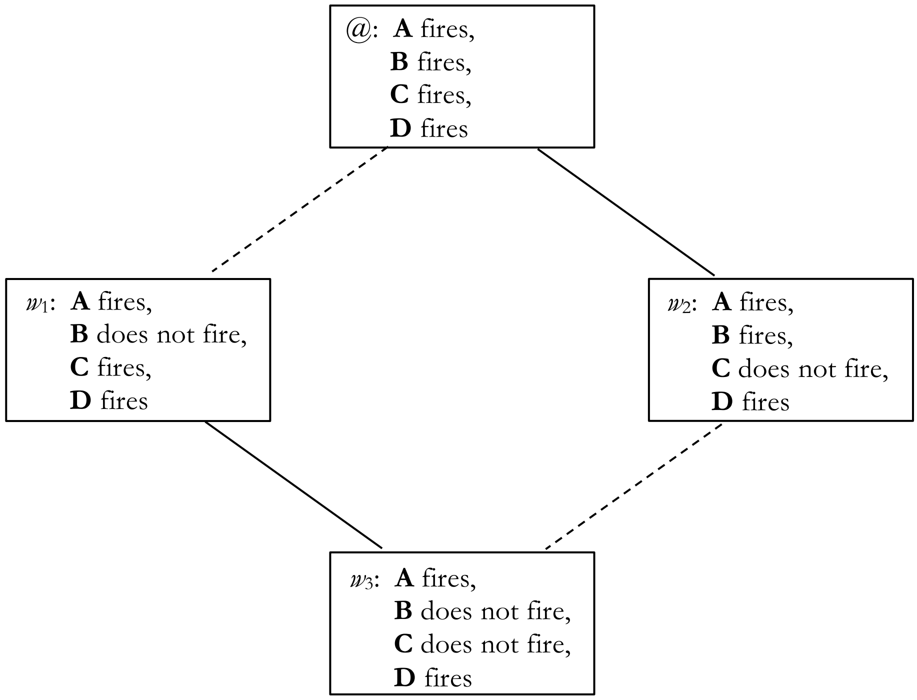
Possibility horizon $\mathscr{H}_1$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *B* cause *J*?	$\mathscr{H}_1$	–	–
Does *C* cause *J*?	$\mathscr{H}_1$	✓	✓
Does *D* cause *J*?	$\mathscr{H}_1$	✓	✓

Note that *A* is here treated as a background condition.

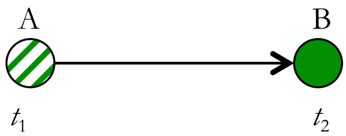
Possibility horizon $\mathscr{H}_2$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *B* cause *J*?	$\mathscr{H}_2$	–	–
Does *C* cause *J*?	$\mathscr{H}_2$	✓	–
Does *D* cause *J*?	$\mathscr{H}_2$	✓	–

Note that *A* and *D* are here treated as background conditions.

15.* Scarlet[15]



Note: **A** can here fire in two different ways – in stripes or in uniform green. As shown in the figure, it in fact fires in stripes. The laws are such that **B** fires if and only if **A** fires in stripes. Let A be the event that is essentially **A**'s firing (but where it is not essential whether **A** fires in stripes or uniform green), and let *A-stripes* be the event that is essentially **A**'s firing in stripes.

A real-life case with this structure:

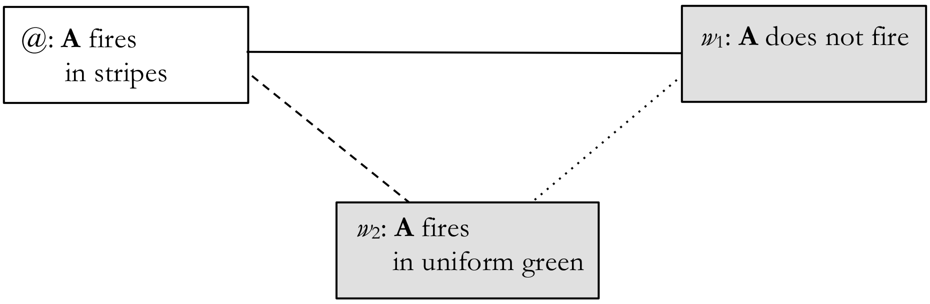
> *Scarlet*: The pigeon Sophia has been conditioned to peck at scarlet to the exclusion of all other colours. She is presented with a scarlet triangle and pecks at it.

A here corresponds to the triangle's being red, *A-stripes* corresponds to the triangle's being scarlet, and B corresponds to Sophia's pecking.

We here consider the effect B and its causes at time t_1. We may consider this case within two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of both of these possibility horizons, worlds in which B does not occur are coloured grey.

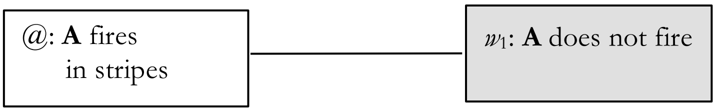
[15] This case is closely based on a case presented in Yablo (1992*a*), p. 257. For similar cases, see Yablo (1992*b*), p. 415, and Sartorio (2010), pp. 266-69.

Possibility horizon $\mathscr{H}_1$:



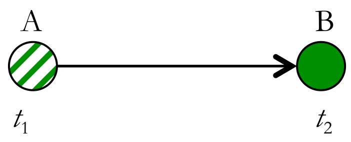
Question	Possibility horizon	Process-connection	Security-dependence
Does *A-stripes* cause *B*?	$\mathscr{H}_1$	✓	✓
Does *A* cause *B*?	$\mathscr{H}_1$	–	✓

Possibility horizon $\mathscr{H}_2$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A-stripes* cause *B*?	$\mathscr{H}_2$	✓	✓
Does *A* cause *B*?	$\mathscr{H}_2$	–	✓

16.* Red[16]



Note: **A** can here fire in two different ways – in stripes or in uniform green. As shown in the figure, it in fact fires in stripes. The laws are such that **B** fires just in case **A** fires (whether **A** fires in stripes or in uniform green). Furthermore, the case is such that in the closest-at-t_1 world(s) where **A** does not fire in stripes, **A** does not fire at all. Let *A* be the event that is essentially **A**'s firing (but where it is not essential whether **A** fires in stripes or uniform green), and let *A-stripes* be the event that is essentially **A**'s firing in stripes.

A real-life case with this structure:

> *Red*: The pigeon Delia has been conditioned to peck at red to the exclusion of all other colours. She is presented with a scarlet triangle and pecks at it. In the lab where she is, the researchers use just two colours – scarlet and emerald. If Delia had not been presented with a scarlet triangle, she would have been presented with an emerald triangle.

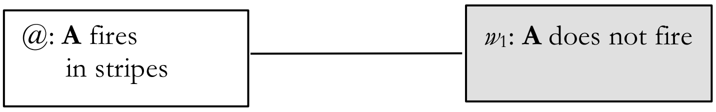
A here corresponds to the triangle's being red, *A-stripes* corresponds to the triangle's being scarlet, and *B* corresponds to Delia's pecking.

We here consider the effect *B* and its causes at time t_1. In my representation of the relevant possibility horizon, worlds in which *B* does not occur are coloured grey.

---

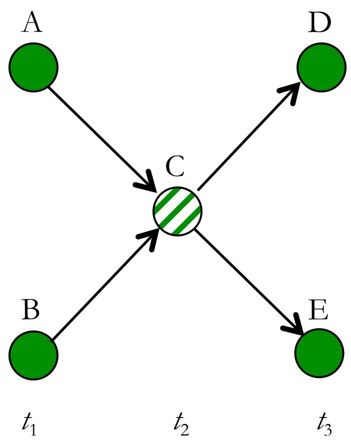
[16] This case is closely based on a case presented in Yablo (1992*a*), p. 257. For similar cases, see Yablo (1992*b*), p. 417, and Sartorio (2010), pp. 264-66.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A-stripes* cause *B*?	$\mathscr{H}$	–	✓
Does *A* cause *B*?	$\mathscr{H}$	✓	✓

17.* Apparent counterexample to transitivity[17]



Note: neuron **C** can here fire in two different ways – in uniform green and in stripes. The neuron laws are such that **C** fires in uniform green if and only if it receives a stimulatory signal from **A** and no stimulatory signal from **B**; and **C** fires in stripes if and only if it receives stimulatory signals from both **A** and **B**. Furthermore, **D** fires if and only if **C** fires (independently of whether **C** fires in uniform green or in stripes). By contrast, **E** is sensitive to the way in which **C** fires – and so, **E** fires if and only if **C** fires in stripes. Let *C* be the event that is essentially **C**'s firing (but where it is not essential whether **C** fires in stripes), and let *C-stripes* be the event that is essentially **C**'s firing in stripes. A real-life case with this structure:

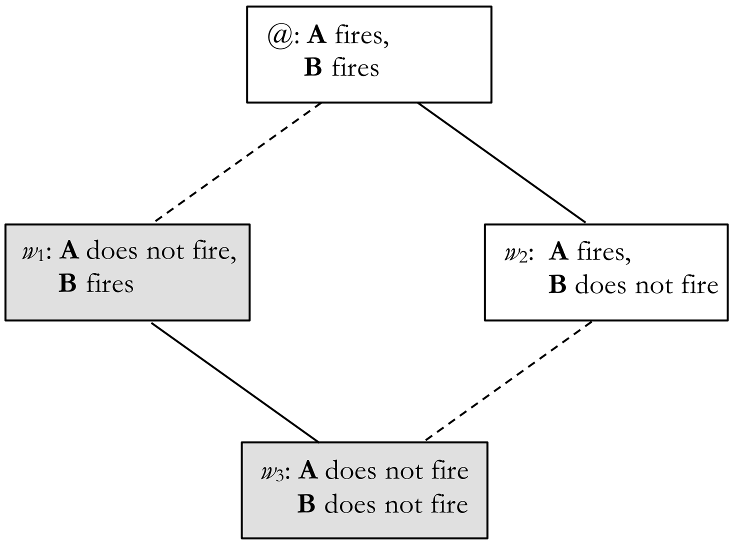
> *Skiing accident*: while skiing, Suzy breaks her right wrist. The next day, she writes a philosophy paper, which is subsequently accepted for publication. Since Suzy's right wrist is broken, she writes the paper by typing with her left hand. As she is not used to writing this way, she develops a cramp in her left hand.

A here corresponds to Suzy's having the idea for her paper, *B* corresponds to the skiing accident, *C* corresponds to Suzy's writing her paper (either normally or by typing with her left hand), *C-stripes* corresponds to Suzy's writing her paper by typing with her left hand, *D* corresponds to the paper's being accepted for publication, and *E* corresponds to Suzy's getting a cramp in her left hand.

[17] Cf. Paul and Hall (2013), Figure 48.

We begin by considering the effect *D* and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which *D* does not happen are coloured grey:

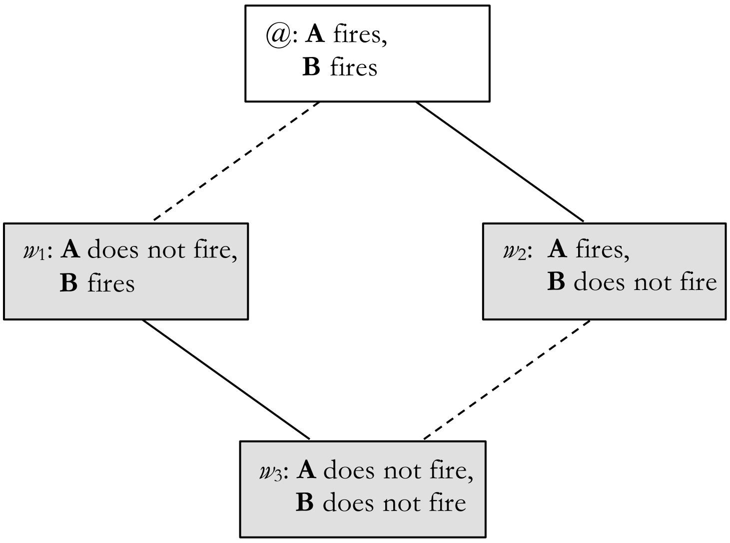
Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *D*?	$\mathscr{H}$	✓	✓
Does *B* cause *D*?	$\mathscr{H}$	–	–

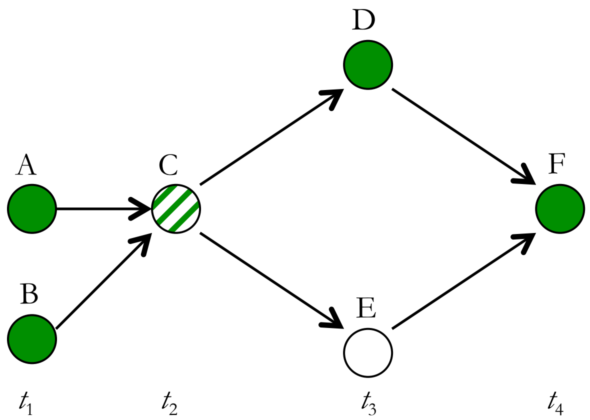
We next consider the effect E and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which E does not happen are coloured grey:

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause E?	$\mathscr{H}$	✓	✓
Does B cause E?	$\mathscr{H}$	✓	✓

18.* Switch[18]



Note: **C** can fire in two different ways – in uniform green and in stripes. **C** fires in uniform green if and only if **A** fires and **B** does not; and **C** fires in stripes if and only if **A** and **B** both fire. If **C** fires in uniform green, it sends a stimulatory signal to **E**, and not to **D**. If **C** fires in stripes, it sends a stimulatory signal to **D**, and not to **E**.

A real-life case with this structure:

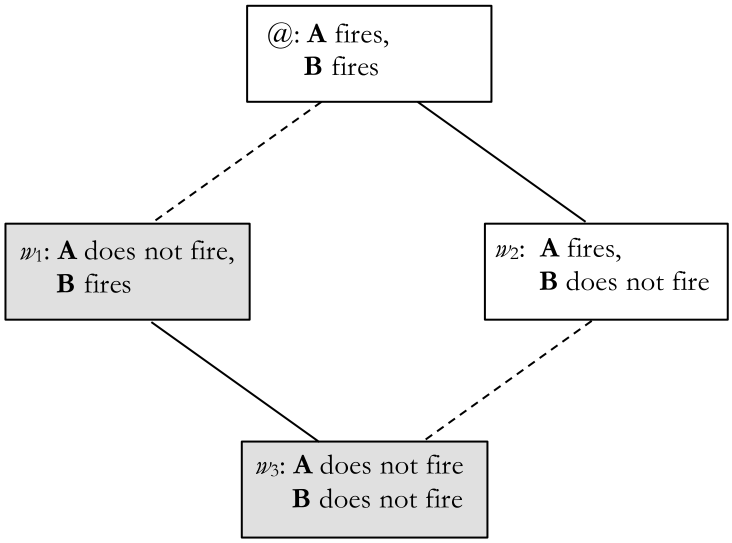
> *Switch*: Suzy is standing by a switch in the railroad tracks. She sees a train approaching in the distance, and flips the switch so that the train travels down the left-hand track. If she had not flipped the switch, the train would have travelled down the right-hand track instead. Since the tracks converge a few miles later, the train arrives at its destination all the same.

A here corresponds to the approaching train, *B* corresponds to Suzy's flipping the switch, *C-stripes* corresponds to the train's being directed towards the left-hand track, and *F* corresponds to the train's arrival.

We here consider the effect *F* and its causes at time t_1 within the possibility horizon illustrated below – where the worlds in which *F* does not happen are coloured grey:

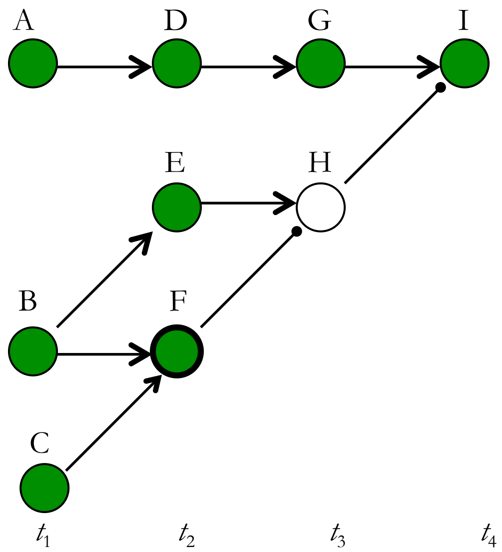
[18] Cf. Paul and Hall (2013), Figure 45.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *I*?	$\mathscr{H}$	✓	✓
Does *B* cause *I*?	$\mathscr{H}$	✓	–

19.* Boulder[19]



Note: **F** is a stubborn neuron that requires two stimulatory signals in order to fire.

A real-life case with this structure:

> *Boulder*: 'A boulder is dislodged and begins rolling ominously toward Hiker. Before it reaches him, Hiker sees the boulder and ducks. The boulder sails harmlessly over his head with nary a centimetre to spare. Hiker survives his ordeal.'

A here corresponds to the physiological processes that keep Hiker alive, *B* corresponds to the boulder's fall, *C* corresponds to Hiker's being attentive, *F* corresponds to Hiker's duck, and *I* corresponds to Hiker's survival.

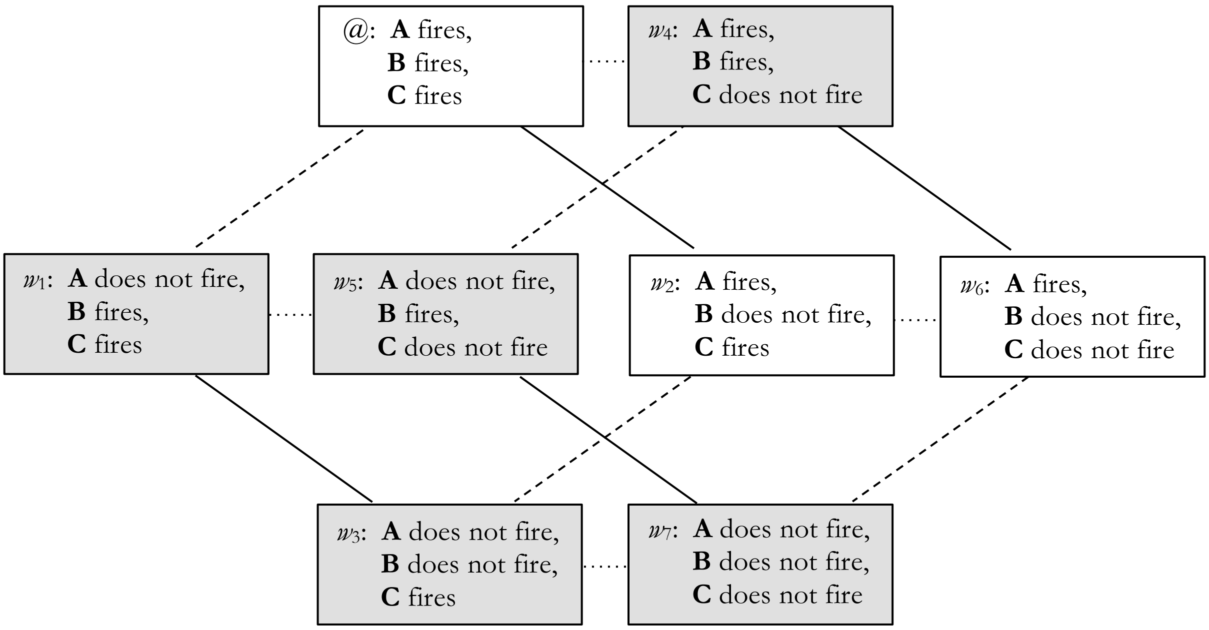
We here consider the effect *I* and its causes at time t_1. We may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of both of these possibility horizons, worlds in which *I* does not happen are coloured grey.

---

[19] This case is a modification of Paul and Hall (2013), Figure 10. Cf. Paul and Hall (2013), p. 222.

Possibility horizon $\mathscr{H}_1$:

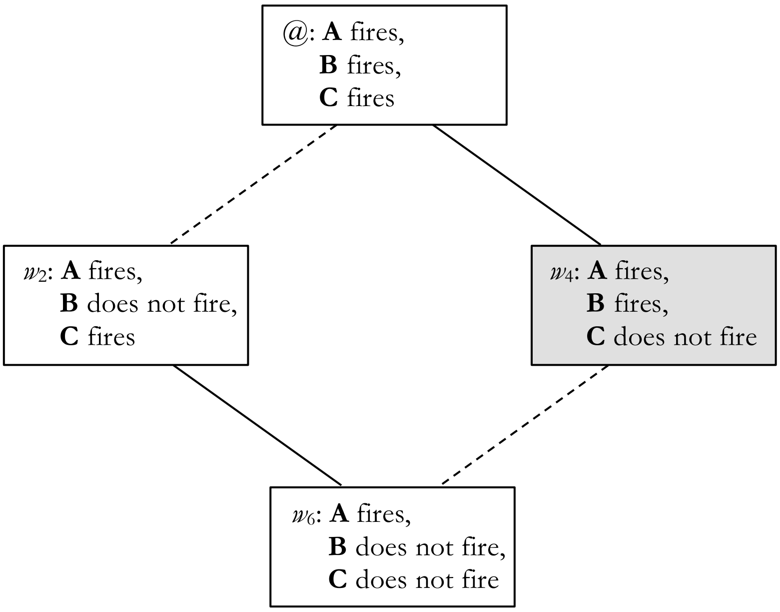
Note: this is not a natural choice of possibility horizon, since we would in most contexts treat *A* (the physiological processes that keep Hiker alive) as a background condition.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *D*?	$\mathscr{H}_1$	✓	✓
Does *B* cause *D*?	$\mathscr{H}_1$	✓	–
Does *C* cause *D*?	$\mathscr{H}_1$	✓	✓

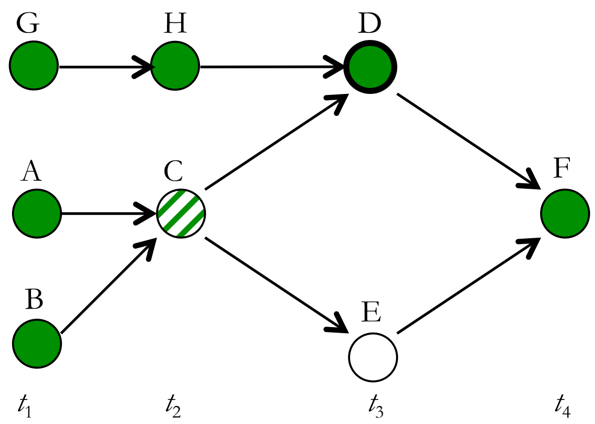
Possibility horizon $\mathscr{H}_2$:

Note: this is a more natural choice of possibility horizon, since it treats *A* (the physiological processes that keep Hiker alive) as a background condition.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *I*?	$\mathscr{H}_2$	✓	–
Does *B* cause *I*?	$\mathscr{H}_2$	✓	–
Does *C* cause *I*?	$\mathscr{H}_2$	✓	✓

20.* Switch – a modified version[20]



Note: the case is as described in *Switch* above, except that **D** is here a stubborn neuron requiring two stimulatory signals in order to fire – one from **H** and one from **C**.

A real-life case with this structure is like *Switch* above, with the following addition:

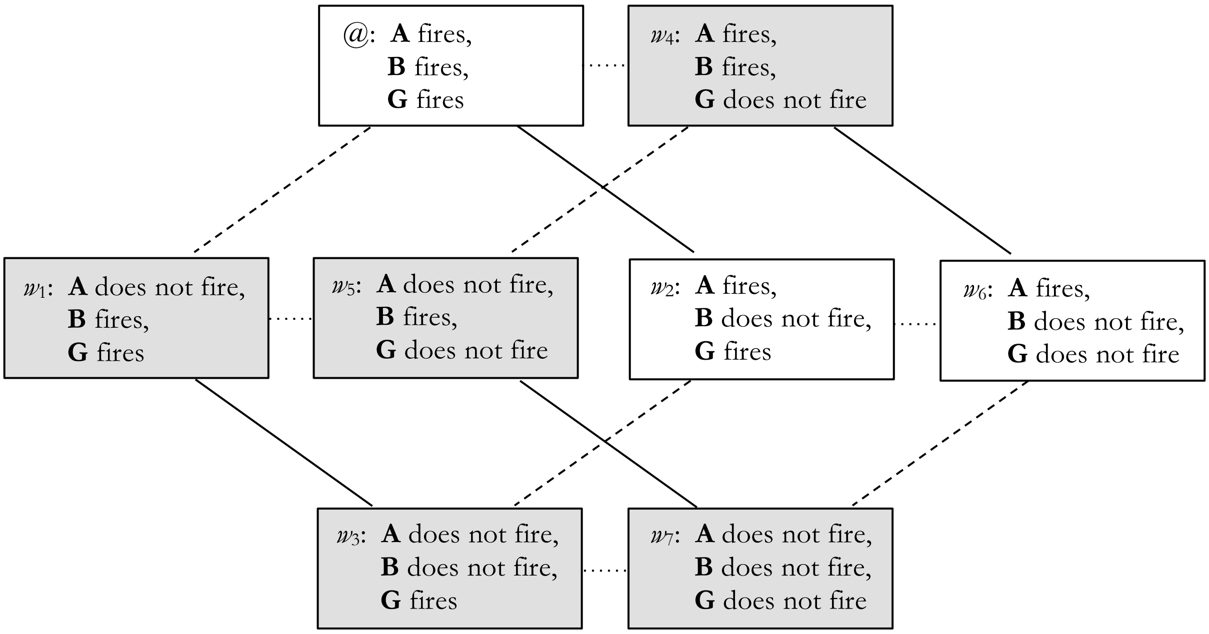
> *Addition to Switch*: When the train is approaching, the left track is in fact disconnected. Just as Suzy flips the switch, Sally comes by and reconnects the track. If Sally had not reconnected the track, the train would have been derailed.

G here corresponds to Sally's reconnecting the left-hand track.

We here consider the effect *F* and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which *F* does not happen are coloured grey:

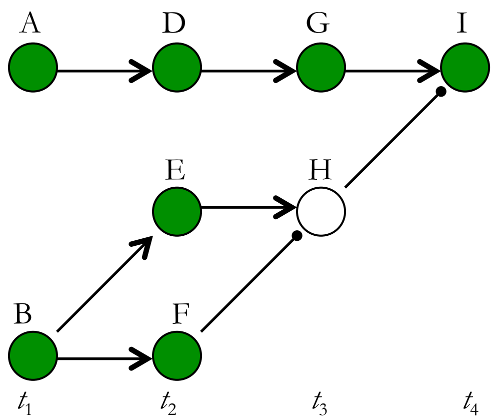
[20] This case is a modified version of Paul and Hall (2013), Figure 45.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *F*?	$\mathscr{H}$	✓	✓
Does *B* cause *F*?	$\mathscr{H}$	✓	–
Does *G* cause *F*?	$\mathscr{H}$	✓	✓

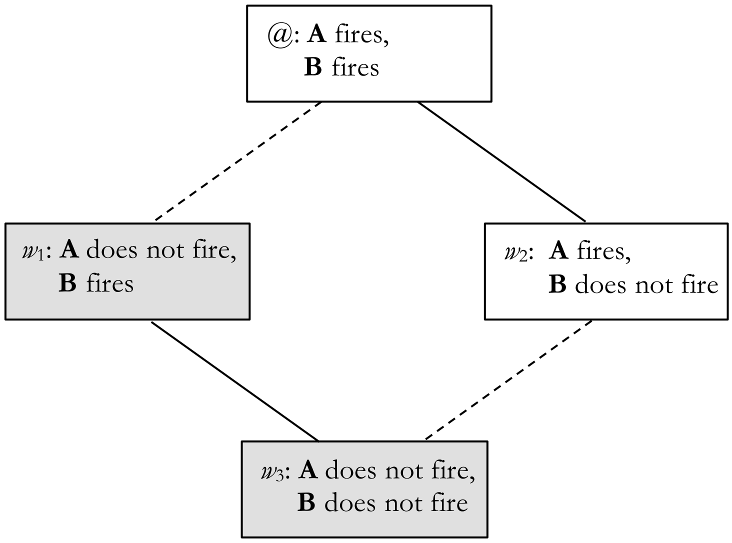
21.* Boulder – a modified version[21]



We here consider the effect *I* and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which *I* does not happen are coloured grey:

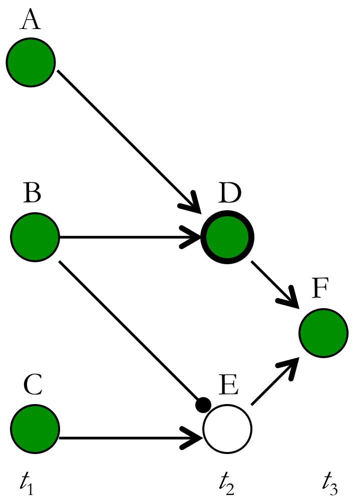
[21] Cf. Paul and Hall (2013), Figure 10.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *I*?	$\mathscr{H}$	✓	✓
Does *B* cause *I*?	$\mathscr{H}$	✓	–

22.* A borderline case[22]



Note: **D** is a stubborn neuron that requires two stimulatory signals to fire.

A real-life case with this structure:

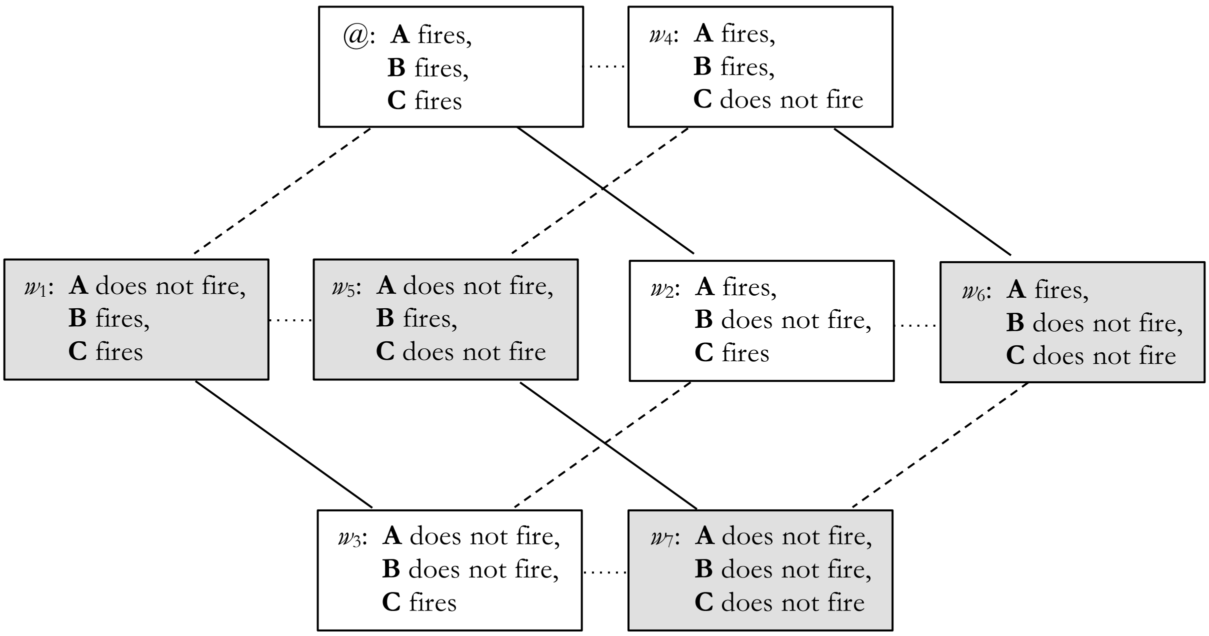
> *Window-shattering*: Suzy throws a rock at the window. Her aim is slightly off, but a strong gust of wind brings her rock back on course. It hits the window and the window shatters. If Suzy had not thrown, Billy would have thrown a larger rock and hit the window, independently of whether there was wind or not.

A here corresponds to the strong gust of wind, *B* corresponds to Suzy's throw, *C* corresponds to Billy's being ready to throw, *D* corresponds to Suzy's rock being on course to hit the window, and *F* corresponds to the window-shattering.

We here consider the effect *F* and its causes at time t_1. We may consider the case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of both of these possibility horizons, worlds where *F* does not occur are coloured grey.

[22] This is a modified version of Paul and Hall (2013), Figure 1.

Possibility horizon $\mathscr{H}_1$:

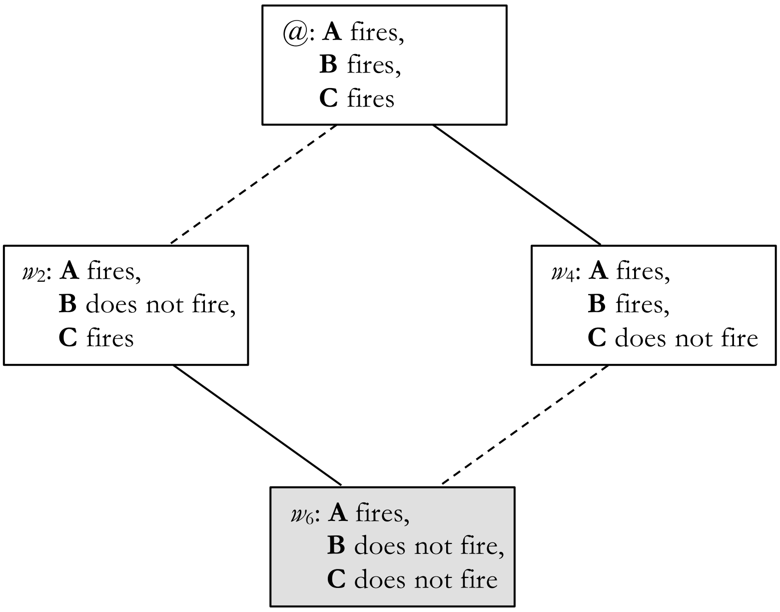


Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *F*?	$\mathscr{H}_1$	✓	✓
Does *B* cause *F*?	$\mathscr{H}_1$	✓	–
Does *C* cause *F*?	$\mathscr{H}_1$	–	✓

The judgements of comparative security are made on the assumption that there is only a short distance-at-t_1 from @ to w_1, where *A* does not occur, but a long distance-at-t_1 from @ to w_4, where *C* does not occur.

Possibility horizon $\mathscr{H}_2$:

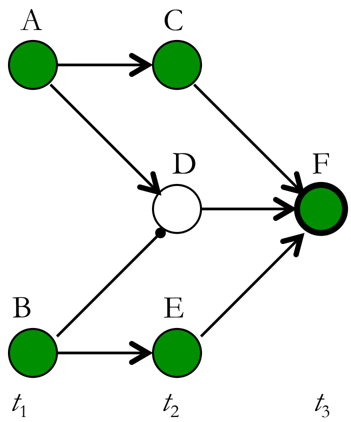
Note: we may artificially restrict our possibility horizon to include only worlds where *A* occurs, by using the phrase 'given that . . .', as in 'given that *A* in fact occurs, *B* is a cause of *F*'. Within this restricted possibility horizon, we find that *F* security-depends on *B*.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *F*?	$\mathscr{H}_2$	✓	–
Does *B* cause *F*?	$\mathscr{H}_2$	✓	✓
Does *C* cause *F*?	$\mathscr{H}_2$	–	✓

Note that *A* is here is treated as a background condition.

23. A complex version of early preemption[23]

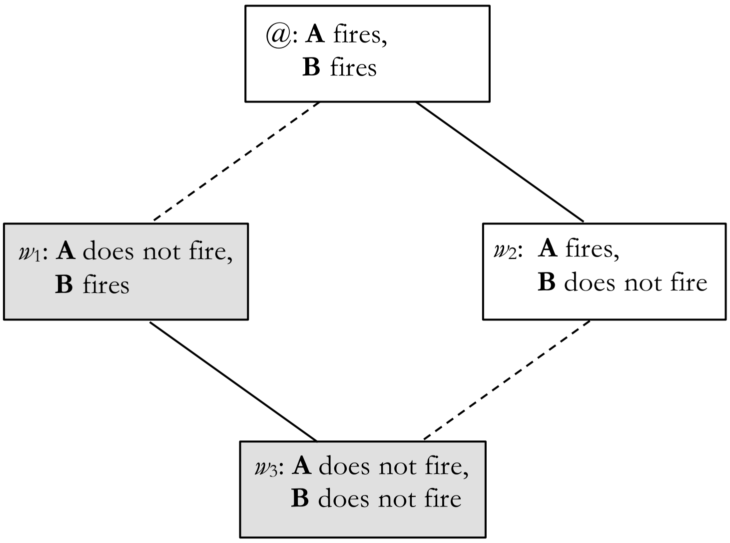


Note: **F** is a stubborn neuron that requires two stimulatory signals in order to fire.

[23] Cf. Paul and Hall (2013), Figure 14.

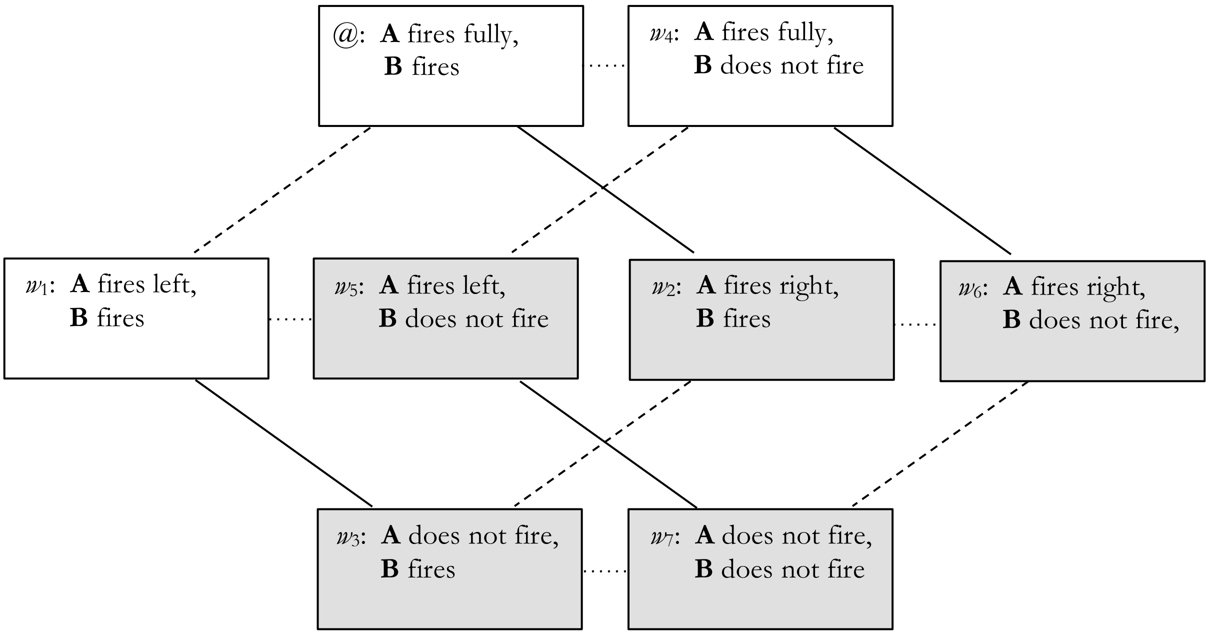
We here consider the effect F and its causes at time t_1. The most natural choice of possibility horizon is $\mathscr{H}_1$, illustrated below:

Possibility horizon $\mathscr{H}_1$:



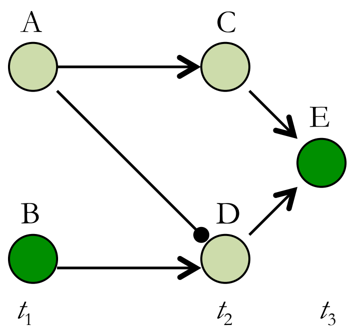
Question	Possibility horizon	Process-connection	Security-dependence
Does A cause F?	$\mathscr{H}_1$	✓	✓
Does B cause F?	$\mathscr{H}_1$	✓	–

Within the $\mathscr{H}_1$, *E* does not security-depend on *B*. However, we can restore *E*'s security-dependence on *B* by distinguishing between three ways in which **A** can fire: firing fully, firing left, and firing right. When **A** fires fully, it sends stimulatory signals to both **C** and **D**; when **A** fires left, it sends a stimulatory signal to **C**, but not to **D**; and when **A** fires right, it sends a stimulatory signal to **D**, but not to **C**. Let *A* be the event that is essentially **A**'s firing fully (we do not allow the disjunctive events of e.g. **A**'s firing either fully or left within the range of our quantifiers). Based on this distinction, we may now consider the possibility horizon $\mathscr{H}_2$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *F*?	$\mathscr{H}_2$	✓	✓
Does *B* cause *F*?	$\mathscr{H}_2$	✓	✓

24. Tampering[24]



Note: in this case, neurons can fire with different intensities. When a neuron is coloured dark green, this indicates that it fires with intensity 2. When a neuron is coloured light green, this indicates that it fires with intensity 1. **A** and **C** can fire in just one way – namely, with intensity 1. **B** and **D** can fire in two ways – with intensity 1 or intensity 2.

The strength with which a neuron fires determines the strength of the stimulatory and inhibitory signals it sends. In the figure, **A** fires with intensity 1. This sends a stimulatory signal of intensity 1 to **C**, which in turn fires with intensity 1. It also sends an inhibitory signal with intensity 1 to **D**. **B** fires with intensity 2 and sends a stimulatory signal of intensity 2 to **D**.

The rule for calculating the intensity with which **D** fires is this: **D** fires with an intensity corresponding to the intensity of the stimulatory signal from **B**, *minus* the intensity of the inhibitory signal from **A**. Thus, **D** fires with intensity 1.

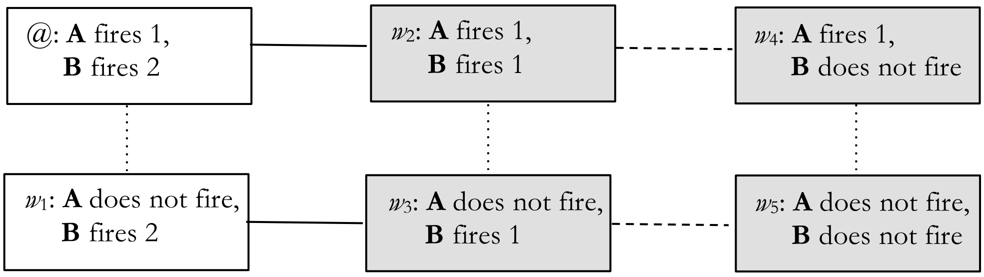
Finally, **E** fires if and only if it receives stimulatory signals of a combined intensity of at least 2. Since it receives a signal of intensity 1 from **C** and a signal of intensity 1 from **D**, **E** fires.

Let *A* be the event that is essentially **A**'s firing; let *B* be the event that is essentially **B**'s firing (but where it is not essential that **B** fires with intensity 2); and let *B-2* be the event that is essentially **B**'s firing with intensity 2.

[24] Cf. Paul and Hall (2013), Figure 15.

We here consider the effect E and its causes at time t_1. We may for example judge this within the possibility horizon $\mathscr{H}$ illustrated below:

Possibility horizon $\mathscr{H}$:



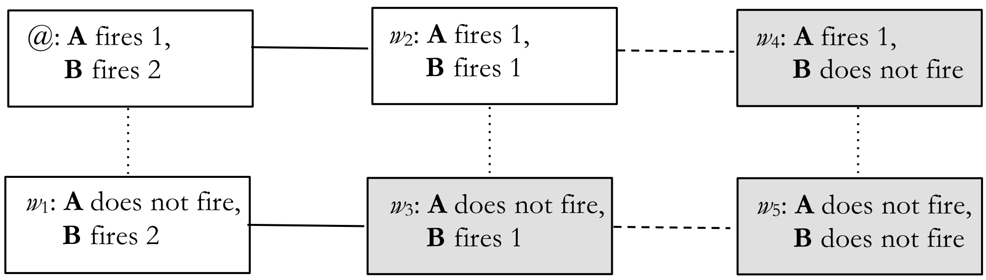
Question	Possibility horizon	Process-connection	Security-dependence
Does A cause E?	$\mathscr{H}$	✓	–
Does B cause E?	$\mathscr{H}$	–	✓
Does B-2 cause E?	$\mathscr{H}$	✓	✓

Note that a slight modification of the rule for calculating the intensity with which **D** fires changes the verdict on whether *A* is a cause of *E*. Consider the following rule:

> If |**A**| = 1 and |**B**| = 2, then |**D**| = 1
> If |**A**| = 1 and |**B**| = 1, then |**D**| = 1
> If |**A**| = 1 and |**B**| = 0, then |**D**| = 0
> If |**A**| = 0 and |**B**| = n, then |**D**| = n.

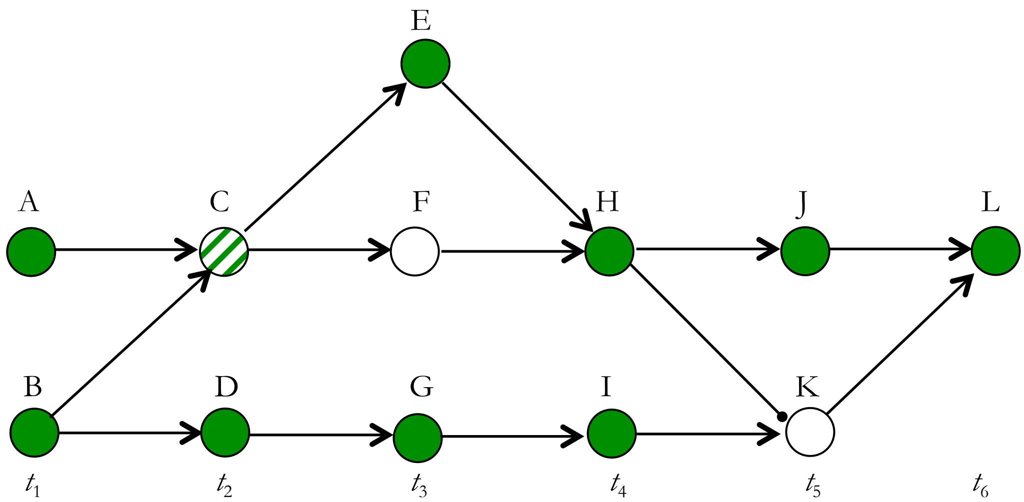
This rule restores *E*'s security-dependence on *A*, within the possibility horizon illustrated below:

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *E*?	$\mathscr{H}$	✓	✓
Does *B* cause *E*?	$\mathscr{H}$	✓	✓
Does *B-2* cause *E*?	$\mathscr{H}$	–	✓

25. Case combining switch and preemption[25]

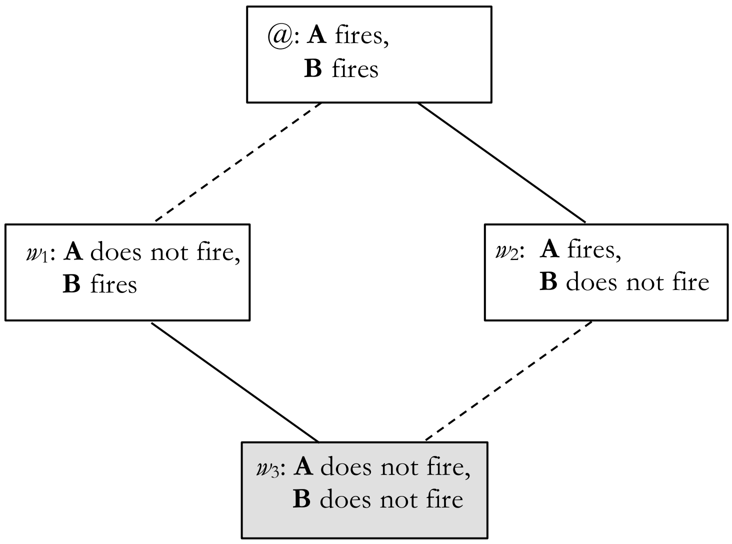


Note: **C** can fire in two different ways – in uniform green or in stripes. **C** fires in uniform green if and only if **A** fires and **B** does not; and **C** fires in stripes if and only if **A** and **B** both fire. If **C** fires in uniform green, it sends a stimulatory signal to **F**, and not to **E**. If **C** fires in stripes, it sends a stimulatory signal to **E**, and not to **F**.

We here consider the effect *L* and its causes at time t_1. We may consider this case based on the possibility horizon illustrated below:

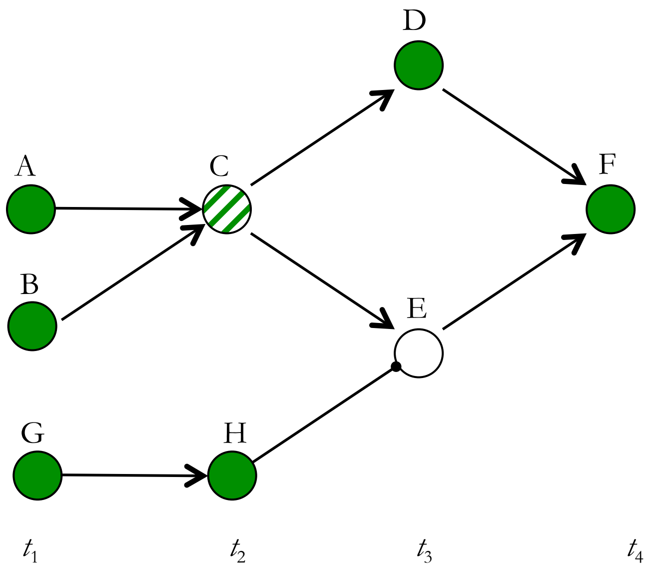
[25] I am grateful to Derek Ball for making me aware of cases of this kind.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *L*?	$\mathscr{H}$	✓	✓
Does *B* cause *L*?	$\mathscr{H}$	✓	✓

26.*   Switch with right-hand track blocked[26]



Note: the case is as described in *Switch* above, except for the addition of neuron **G** and **H**.

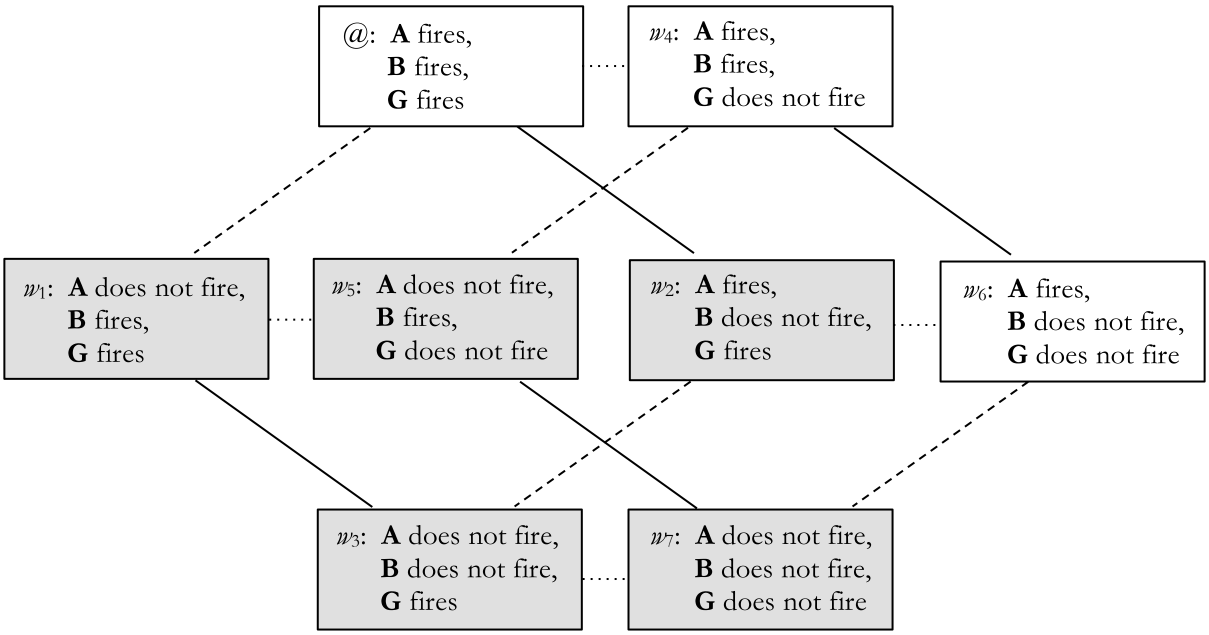
A real-life case with this structure is like *Switch* above, with the addition that the right-hand track is blocked by a fallen tree. *G* here represents the presence of the tree lying across the right-hand track.

We here consider the effect *F* and its causes at time t_1. We may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of these possibility horizons, the worlds where *F* does not occur are coloured grey.

[26] This is a modified version of Paul and Hall (2013), Figure 45, and makes the same point as Paul and Hall (2013), Figure 47.

Possibility horizon $\mathscr{H}_1$:

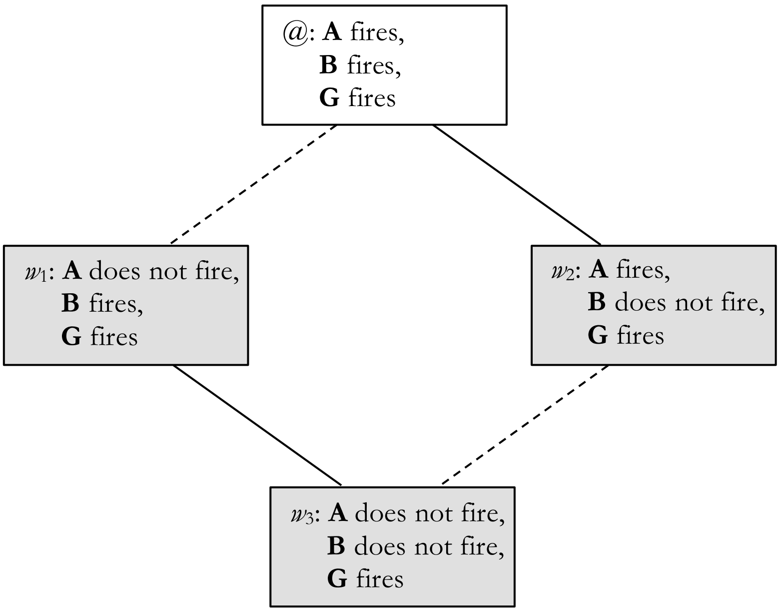
Note: this is a reasonable choice of possibility horizon, treating all of *A*, *B*, and *G* as candidate causes.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *F*?	$\mathscr{H}_1$	✓	✓
Does *B* cause *F*?	$\mathscr{H}_1$	✓	✓
Does *G* cause *F*?	$\mathscr{H}_1$	–	–

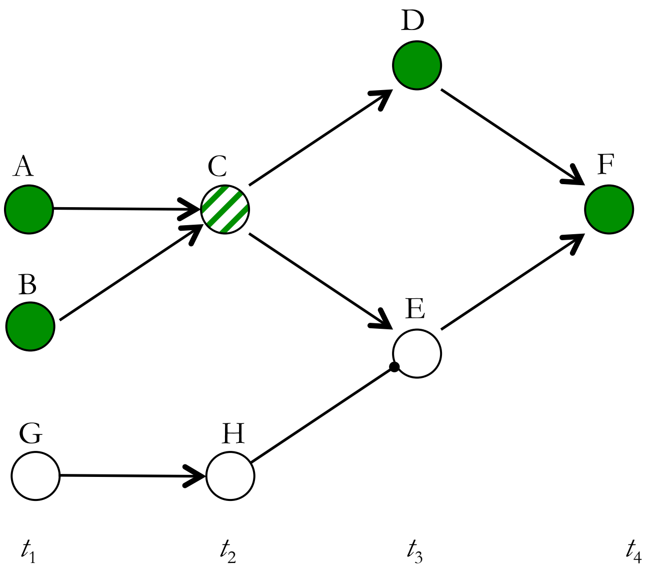
Possibility horizon $\mathscr{H}_2$:

Note: this is also a reasonable choice of possibility horizon, treating *G* (the presence of the tree lying across the right-hand track) as a background condition.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *F*?	$\mathscr{H}_2$	✓	✓
Does *B* cause *F*?	$\mathscr{H}_2$	✓	✓
Does *G* cause *F*?	$\mathscr{H}_2$	–	–

27.* Switch – with right-hand track clear[27]



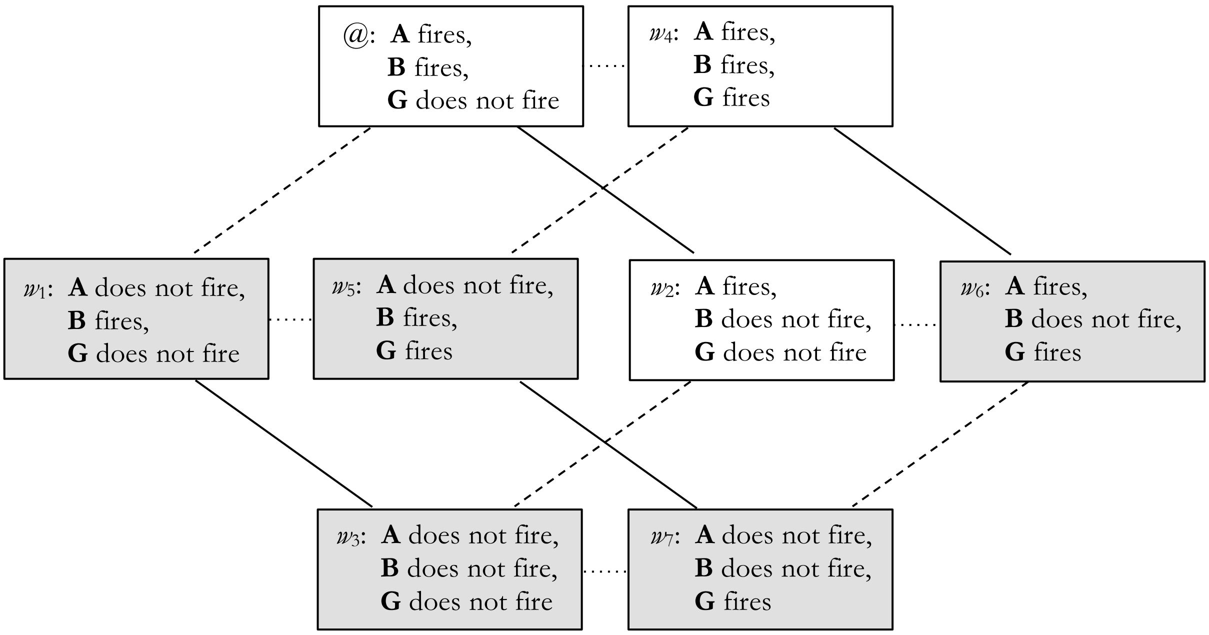
This case is exactly like *Switch with right-hand track blocked*, except that neuron **G** does not fire.

We here consider the effect F and its causes at time t_1. We may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of both of these possibility horizons, worlds in which F does not happen are coloured grey:

[27] This is a modified version of Paul and Hall (2013), Figure 45.

Possibility horizon $\mathscr{H}_1$:

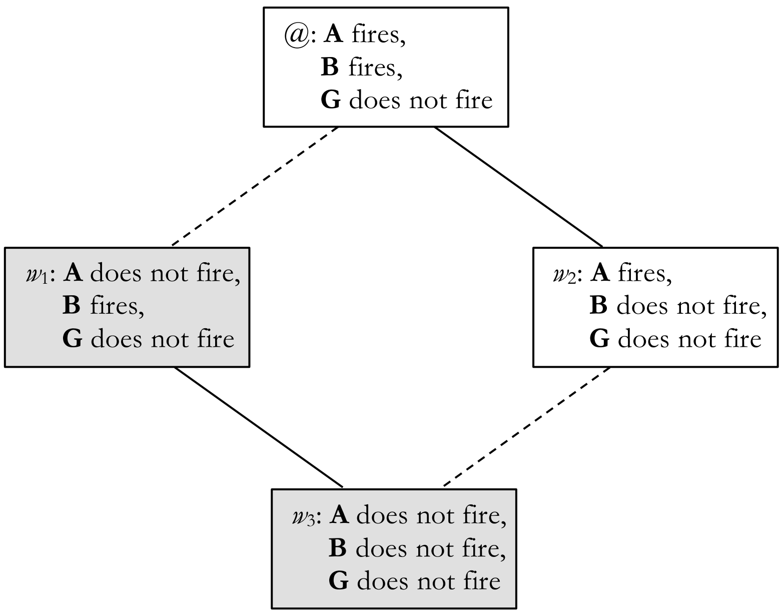
Note: this is not a natural choice of possibility horizon, sine we will in most contexts treat $\neg G$ (the absence of a tree lying across the right-hand track) as a background condition.



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause F?	$\mathscr{H}_1$	✓	✓
Does B cause F?	$\mathscr{H}_1$	✓	✓
Does $\neg G$ cause F?	$\mathscr{H}_1$	–	✓

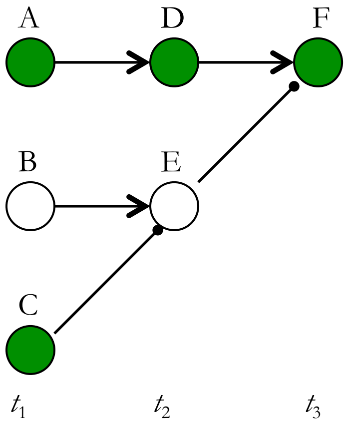
Possibility horizon $\mathscr{H}_2$:

Note: this is a more natural choice of possibility horizon, since it treats $\neg G$ (the absence of a fallen tree lying across the right-hand track) as a background condition.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *F*?	$\mathscr{H}_2$	✓	✓
Does *B* cause *F*?	$\mathscr{H}_2$	✓	–
Does $\neg G$ cause *F*?	$\mathscr{H}_2$	–	–

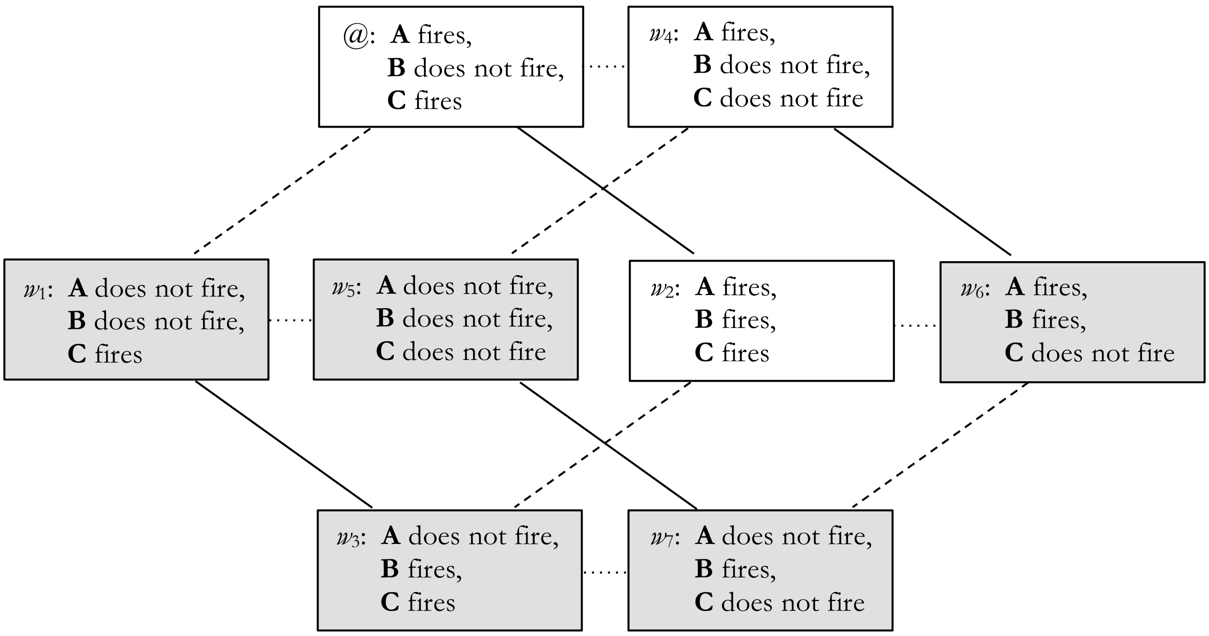
28.*   Double prevention – without threat[28]



We here consider the effect F and its causes at time t_1. We may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of both of these possibility horizons, worlds in which F does not happen are coloured grey:

[28] This is a modified version of Paul and Hall (2013), Figure 29, and makes the same point as Paul and Hall (2013), Figure 40.

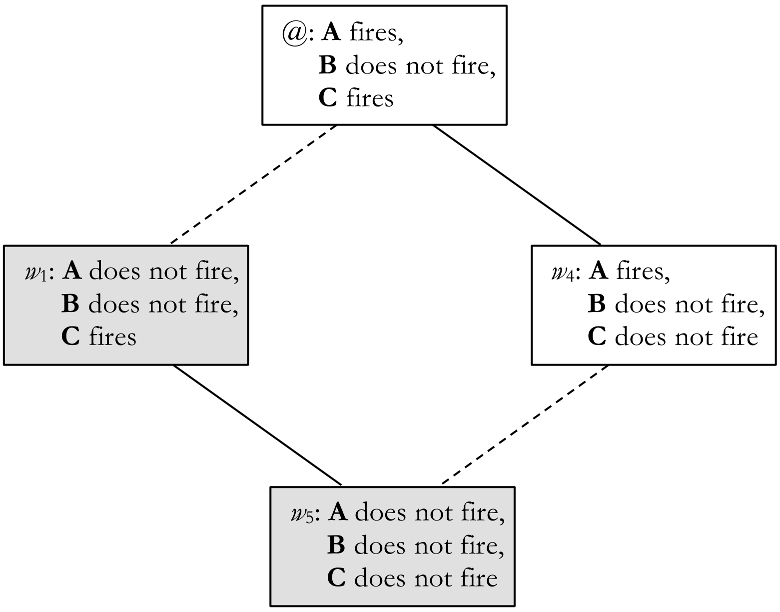
Possibility horizon $\mathscr{H}_1$:



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause F?	$\mathscr{H}_1$	✓	✓
Does $\neg B$ cause F?	$\mathscr{H}_1$	✓	✓
Does C cause F?	$\mathscr{H}_1$	✓	✓

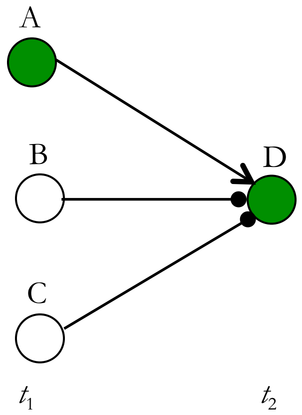
Possibility horizon $\mathscr{H}_2$:

Note: this more restricted possibility horizon, where $\neg B$ is treated as a background condition, is just as reasonable as possibility horizon $\mathscr{H}_1$.



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause F?	$\mathscr{H}_2$	✓	✓
Does $\neg B$ cause F?	$\mathscr{H}_2$	✓	–
Does C cause F?	$\mathscr{H}_2$	✓	–

29.* The queen and the flowers[29]



A real-life case with this structure:

> *The flowers*: Suzy goes on holiday and Billy promises to water her flowers while she is away. However, Billy does not water the flowers and the flowers die.

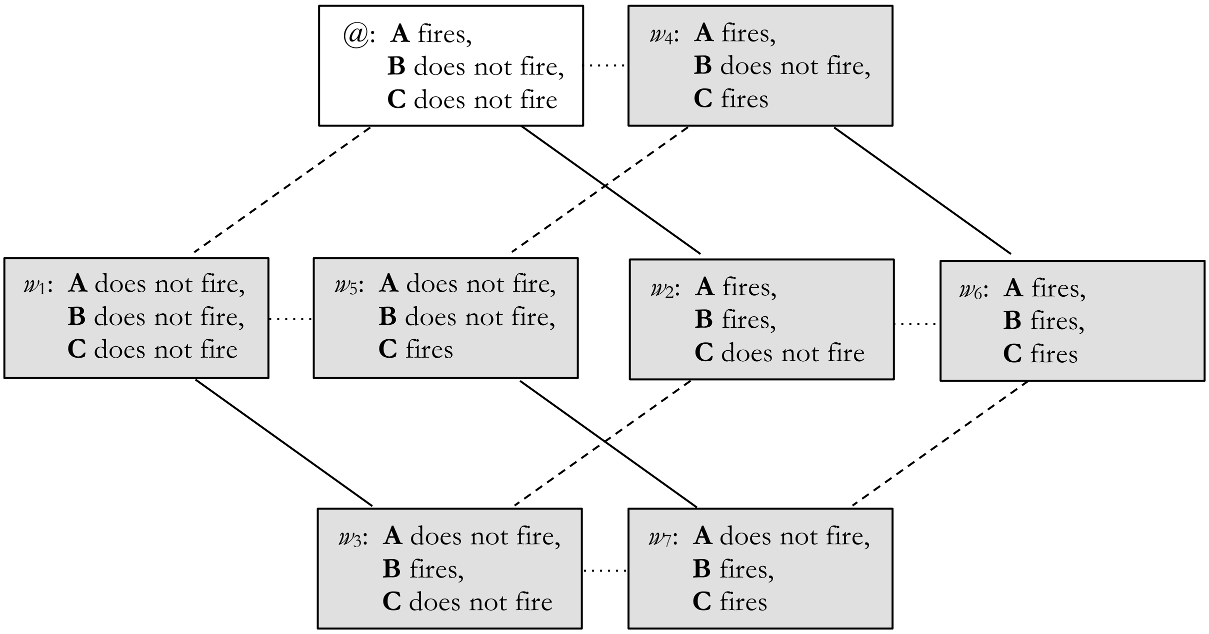
The queen also does not water the flowers. *A* here represents the dry soil in the pot, $\neg B$ represents Billy's failure to water the flowers, $\neg C$ represents the queen's failure to water the flowers, and *D* represents the death of the flowers.

We here consider the effect *D* and its causes at time t_1. We may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of both of these possibility horizons, the worlds where *D* does not occur are coloured grey.

[29] This figure is closely based on a case discussed e.g. in Schaffer (2005), p. 300.

Possibility horizon $\mathscr{H}_1$:

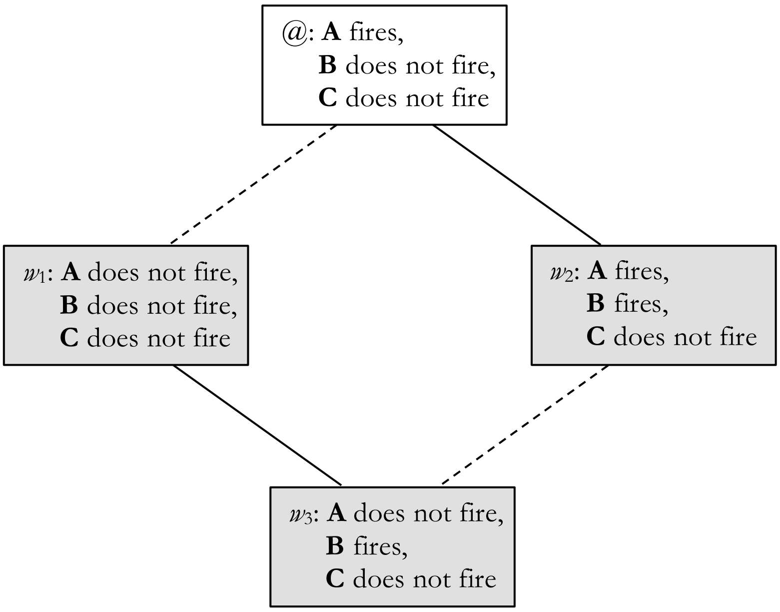
Note: this is not a natural choice of possibility horizon, since we would in most contexts treat ¬*C* (the queen's failure to water the flowers) as a background condition.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *D*?	$\mathscr{H}_1$	✓	✓
Does ¬*B* cause *D*?	$\mathscr{H}_1$	✓	✓
Does ¬*C* cause *D*?	$\mathscr{H}_1$	✓	✓

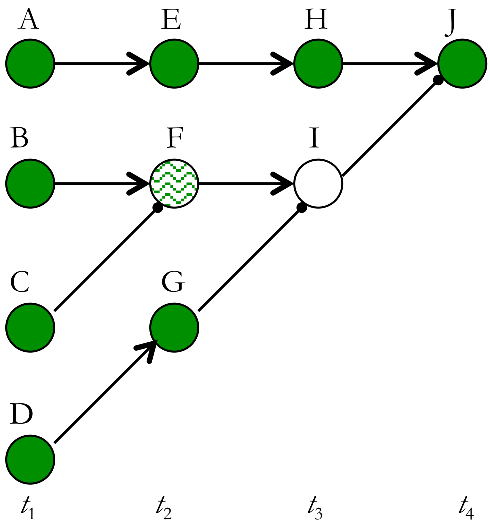
Possibility horizon $\mathscr{H}_2$:

Note: this is a more natural choice of possibility horizon, since it treats $\neg C$ (the queen's failure to water the flowers) as a background condition.



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause D?	$\mathscr{H}_2$	✓	✓
Does $\neg B$ cause D?	$\mathscr{H}_2$	✓	✓
Does $\neg C$ cause D?	$\mathscr{H}_2$	✓	–

30.* McDermott's case[30]



Note: in this case, a neuron has two states in which it does not send any outgoing signals – it can be dormant or it can fire in waves. A neuron is dormant if and only if it receives no stimulatory signals; a neuron fires in waves if and only if it receives at least one stimulatory signal and at least one inhibitory signal. Apart from this, the neuron laws are as usual. The following two cases both have this structure:

> *Wall and window*: Suzy throws a ball towards a window. Before the ball reaches the window, Billy leaps up and catches the ball. If Billy had not caught the ball, it still would not have hit the window – for between Billy and the window, there is a sturdy brick wall.

> *Catcher and window*: Suzy throws a ball towards a window. Before the ball reaches the window, Billy leaps up and catches the ball. If Billy had not caught the ball, it still would not have hit the window – for between Billy and the window stands Sally, who would have caught the ball if Billy had failed to do so.

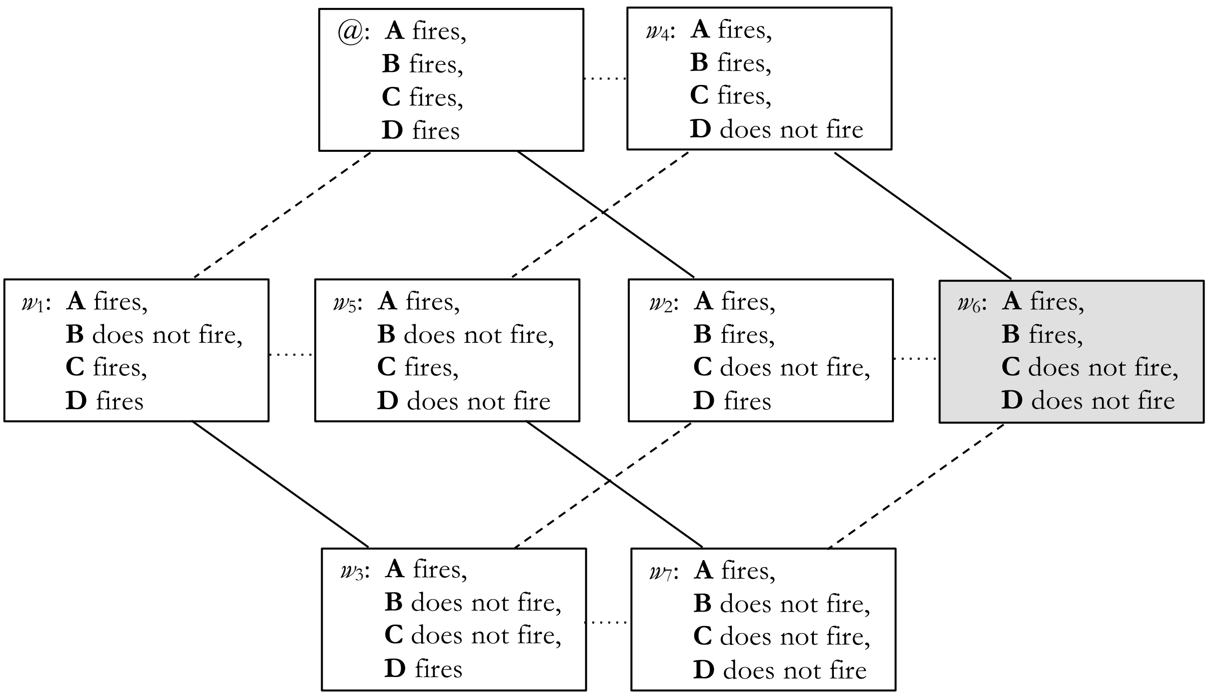
A represents the presence of the window, *B* represents Suzy's throw, *C* represents Billy's being ready to catch the ball, *D* represents the presence of the wall (or Sally's readiness to catch the ball), and *J* represents the window's remaining intact.

[30] This case was originally presented in McDermott (1995), p. 525.

We here consider the effect *J* and its causes at time t_1. Treating *A* as a default event, we may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$.

Possibility horizon $\mathscr{H}_1$:

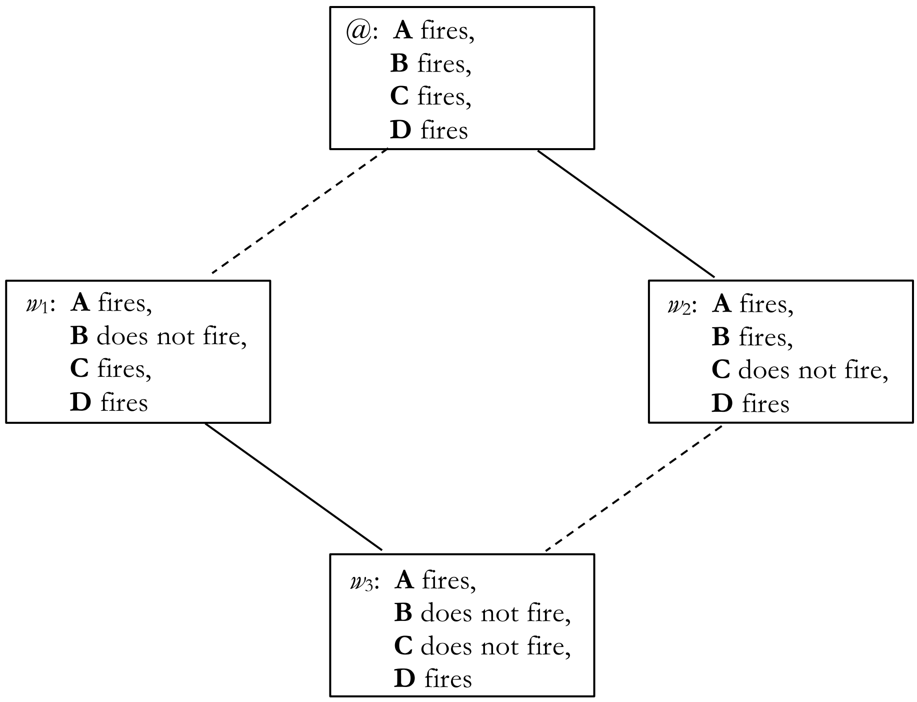
Note: this possibility horizon treats *D* as a candidate cause. This is a natural choice in the case where *D* represents Sally's readiness to catch the ball, since this is categorised as a deviant event in most contexts. By contrast, it is not a natural choice in the case where *D* represents the presence of the wall, since this is categorised as a default event in most contexts.



Question	Possibility horizon	Process-connection	Security-dependence
Does *B* cause *J*?	$\mathscr{H}_1$	✓	–
Does *C* cause *J*?	$\mathscr{H}_1$	✓	✓
Does *D* cause *J*?	$\mathscr{H}_1$	–	✓

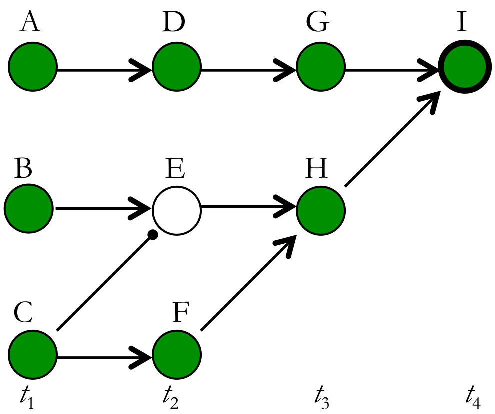
Possibility horizon $\mathscr{H}_2$:

Note: This possibility horizon treats *D* as a background condition. This is a natural choice in the case where *D* represents the presence of the wall, since this is categorised as a default event in most contexts. By contrast, it is not a natural choice in the case where *D* represents Sally's readiness to catch the ball, since this is categorised as a deviant event in most contexts.



Question	Possibility horizon	Process-connection	Security-dependence
Does *B* cause *J*?	$\mathscr{H}_2$	✓	–
Does *C* cause *J*?	$\mathscr{H}_2$	✓	–
Does *D* cause *J*?	$\mathscr{H}_2$	–	–

31. First of Hall's two structurally isomorphic cases[31]

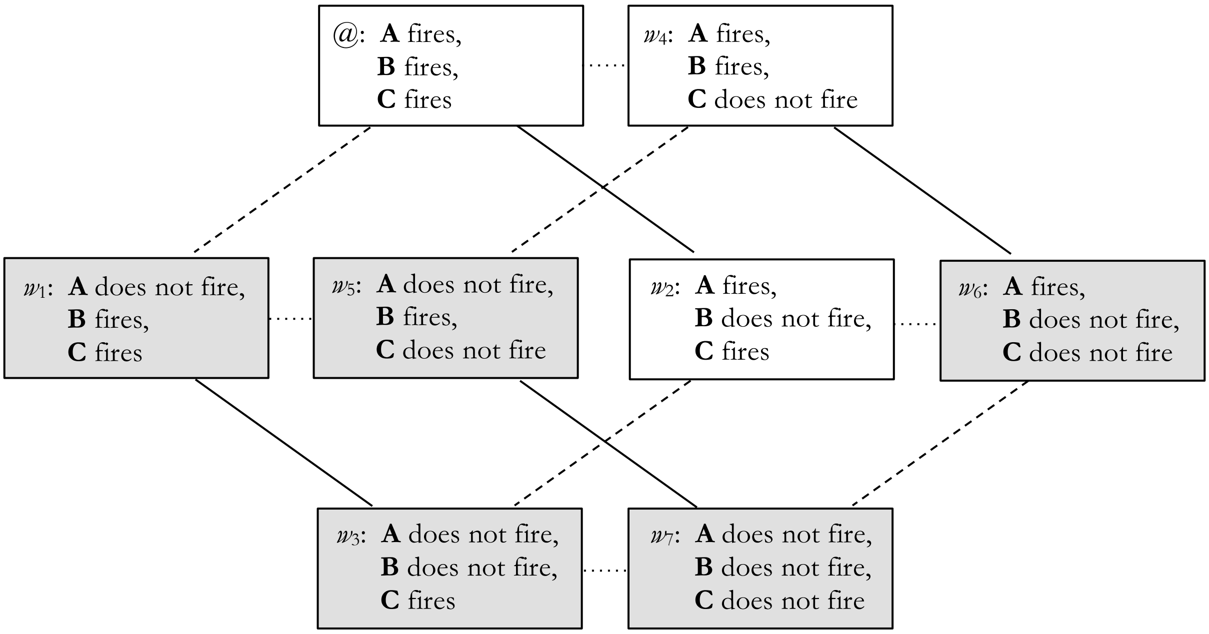


Note: neuron **I** is here a stubborn neuron that requires two stimulatory signals in order to fire.

We here consider the effect *I* and its causes at time t_1. We may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of both of these possibility horizons, the worlds in which *I* does not occur are coloured grey.

[31] Cf. Hall (2007*b*), p. 44; Paul and Hall (2013), Figure 42.

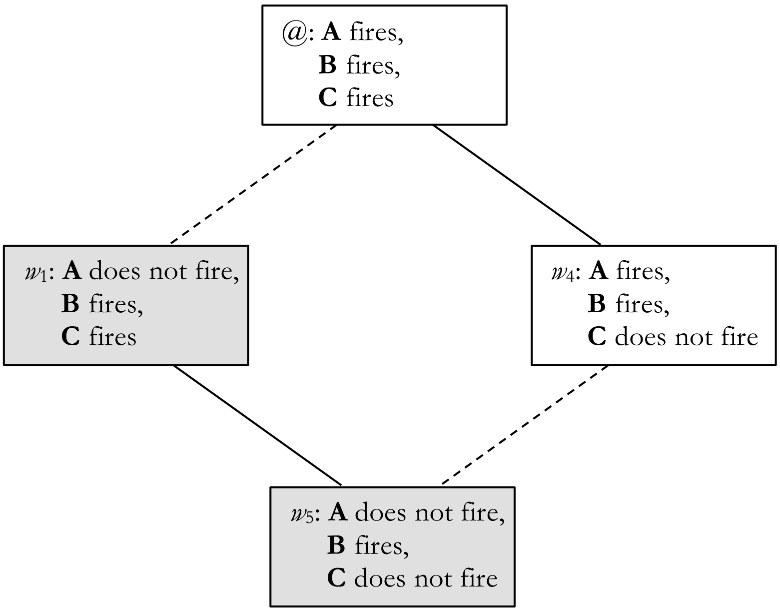
Possibility horizon $\mathscr{H}_1$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *I*?	$\mathscr{H}_1$	✓	✓
Does *B* cause *I*?	$\mathscr{H}_1$	–	✓
Does *C* cause *I*?	$\mathscr{H}_1$	✓	✓

Possibility horizon $\mathscr{H}_2$:

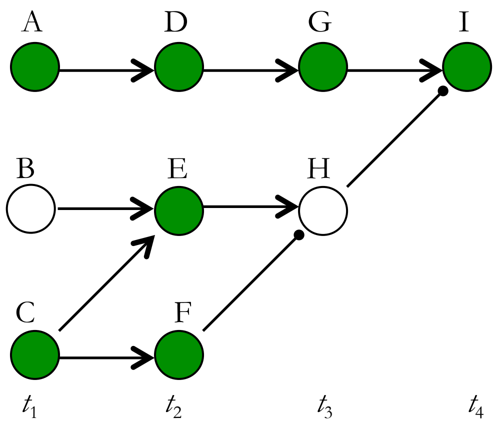
Note: This possibility horizon treats *B* as a background condition. This is not a natural choice, since it goes against the standard practice of treating neuron firings as deviant events when we are dealing with un-interpreted neuron diagrams.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *I*?	$\mathscr{H}_2$	✓	✓
Does *B* cause *I*?	$\mathscr{H}_2$	–	–
Does *C* cause *I*?	$\mathscr{H}_2$	✓	–

Note that *B* is here treated as a background condition.

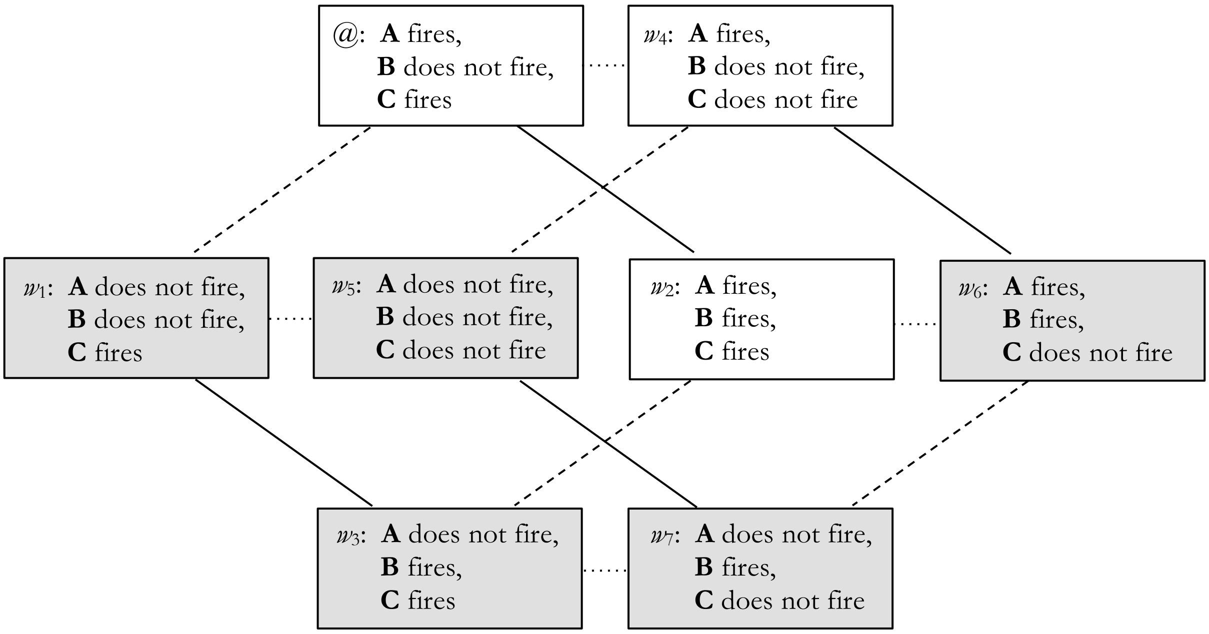
32. Second of Hall's two structurally isomorphic cases[32]



We here consider the effect *I* and its causes at time t_1. We may consider this case based on two different possibility horizons, $\mathscr{H}_1$ and $\mathscr{H}_2$. In my representation of both of these possibility horizons, the worlds in which *I* does not occur are coloured grey.

[32] Cf. Hall (2007*b*), p. 43; Paul and Hall (2013), Figure 41.

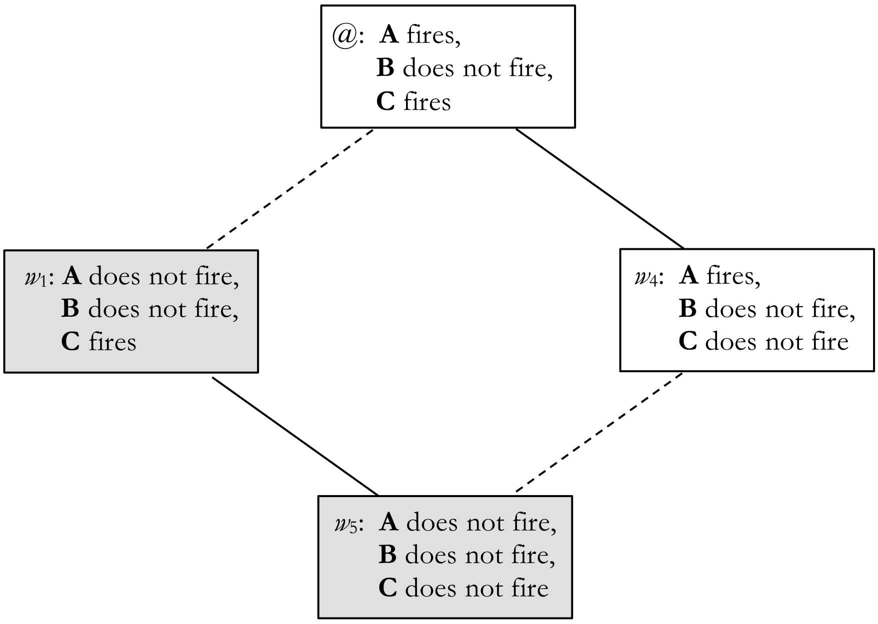
Possibility horizon $\mathscr{H}_1$:



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause I?	$\mathscr{H}_1$	✓	✓
Does $\neg B$ cause I?	$\mathscr{H}_1$	–	✓
Does C cause I?	$\mathscr{H}_1$	✓	✓

Possibility horizon $\mathscr{H}_2$:

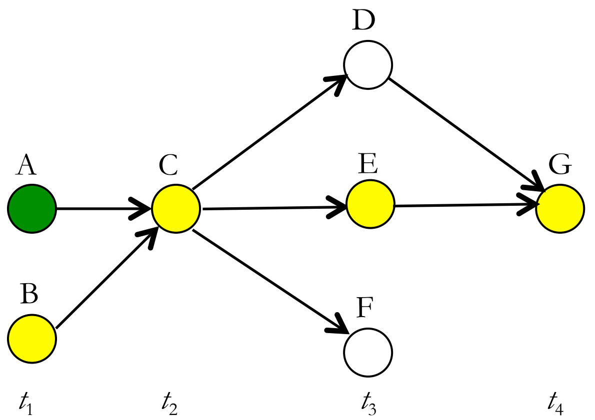
Note: in this possibility horizon, $\neg B$ is treated as a background condition. This possibility horizon is just as appropriate as $\mathscr{H}_1$, since it is appropriate to treat a neuron's failure to fire as a default event when we are dealing within un-interpreted neuron diagrams.



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *I*?	$\mathscr{H}_2$	✓	✓
Does $\neg B$ cause *I*?	$\mathscr{H}_2$	–	–
Does *C* cause *I*?	$\mathscr{H}_2$	✓	–

Note that $\neg B$ is here treated as a background condition.

33.* Contrastive causation[33]



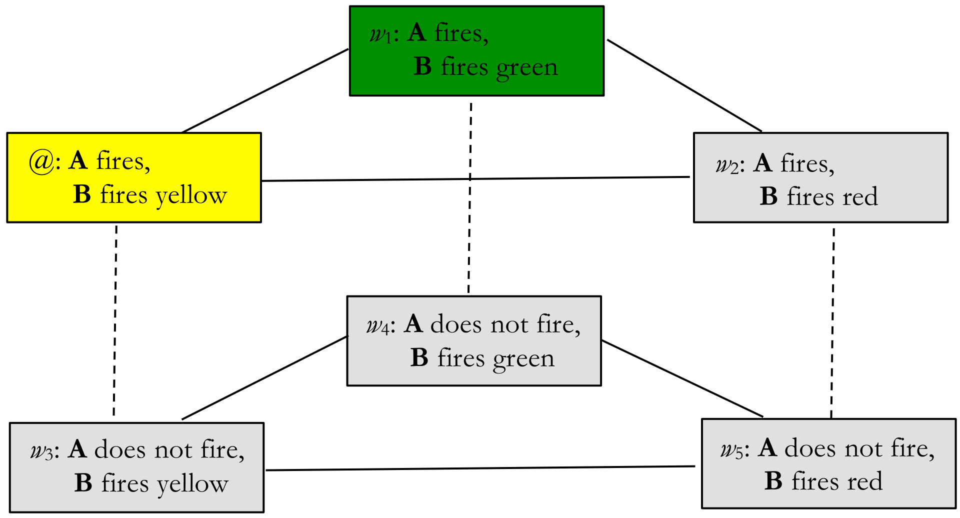
Neuron **B** can here be in three different states: it can fire in green, yellow, or red (it cannot be dormant). Neuron **C** can be in four different states: it can fire in green, yellow, or red, or it can be dormant. **C** fires in green if and only if **A** fires and **B** fires in green; **C** fires in yellow if and only if **A** fires and **B** fires in yellow; and **C** fires in red if and only if **A** fires and **B** fires in red. Finally, **C** is dormant if and only if **A** does not fire. The neuron laws are such that when **C** fires in green, it sends a stimulatory signal to **D** (and no signal to **E** or **F**); when **C** fires in yellow, it sends a stimulatory signal to **E** (and no signal to **D** or **F**); and when **C** fires in red, it sends a stimulatory signal to **F** (and no signal to **D** or **E**). Finally, **G** fires in green if and only if it receives a stimulatory signal from **D**; **G** fires in yellow if and only if it receives a stimulatory signal from **E**; and otherwise **G** remains dormant.

A real-life case with this structure:

> *Switch - variation*: Suzy is standing by a switch in the railroad tracks. She sees a train approaching in the distance, and flips the switch so that the train travels down the local track. She could also have flipped the switch so that the train had travelled down the express track, or the broken track. If she had chosen the express track, the train would have arrived at its destination more quickly. If she had chosen the broken track, the train would have derailed.

33 A case with this structure is presented in Schaffer (2012), pp. 38-39.

The possibility horizon $\mathscr{H}_1$ below includes all the nomologically possible worlds in this case. The worlds in which *G* occurs match the colour in which **G** fires; the worlds in which *G* does not occur are coloured grey:

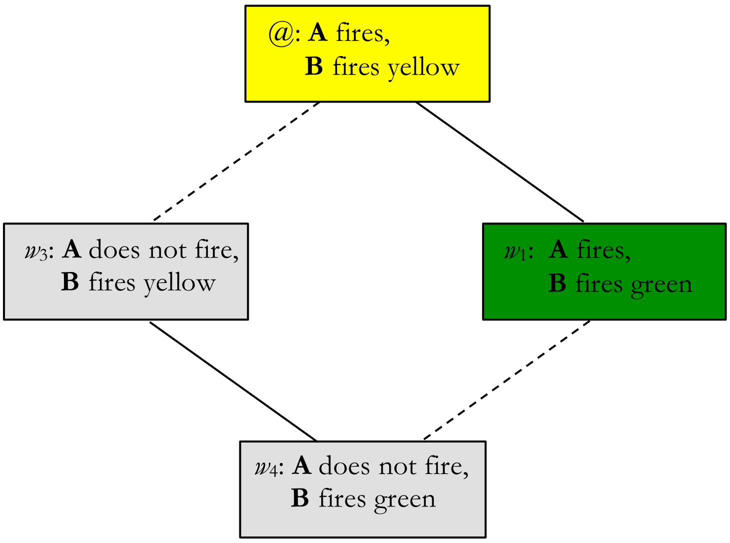


Note that I am assuming that the distance-at-t_1 between @ and w_1 is exactly the same as the distance-at-t_1 between @ and w_2, which is exactly the same as the distance-at-t_1 between w_1 and w_2.

In the following I will illustrate how we may handle a few sample contrastive causal claims.

Consider first the contrastive claim '*B-yellow* rather than *B-green* is a cause of *G*'. To evaluate this claim, we need to consider the restricted possibility horizon $\mathscr{H}_2$, where either *B-green* or *B-yellow* occurs in every world:

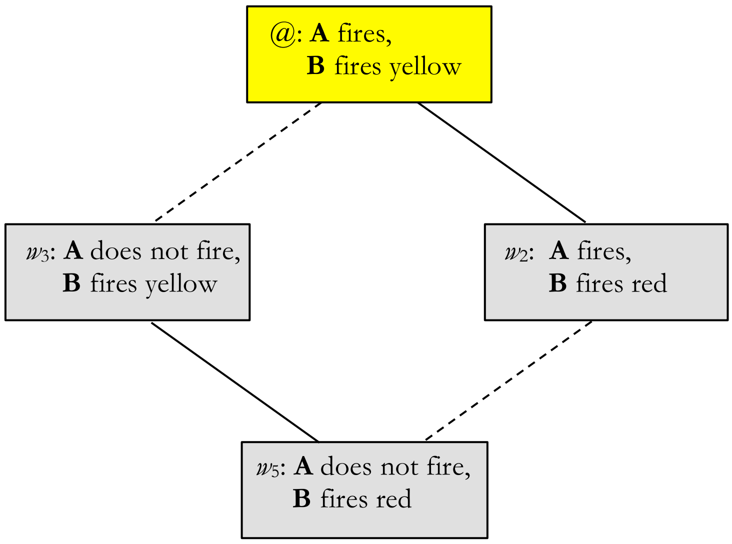
Possibility horizon $\mathscr{H}_2$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *B-yellow* rather than *B-green* cause *G*?	$\mathscr{H}_2$	✓	–

Next, consider the contrastive claim '*B-yellow* rather than *B-red* is a cause of *G*'. To evaluate this claim, we need to consider the restricted possibility horizon $\mathscr{H}_3$ illustrated below:

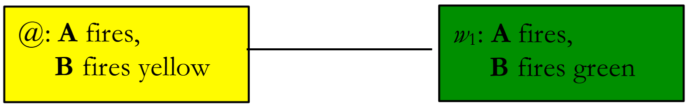
Possibility horizon $\mathscr{H}_3$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *B-yellow* rather than *B-red* cause *G*?	$\mathscr{H}_3$	✓	✓

As our next example, consider the two contrastive claims '*A* caused *G-yellow* rather than *G-green*' and '*B-yellow* caused *G-yellow* rather than *G-green*'. To evaluate both of these claims, we need to consider the restricted possibility horizon $\mathscr{H}_4$ illustrated below:

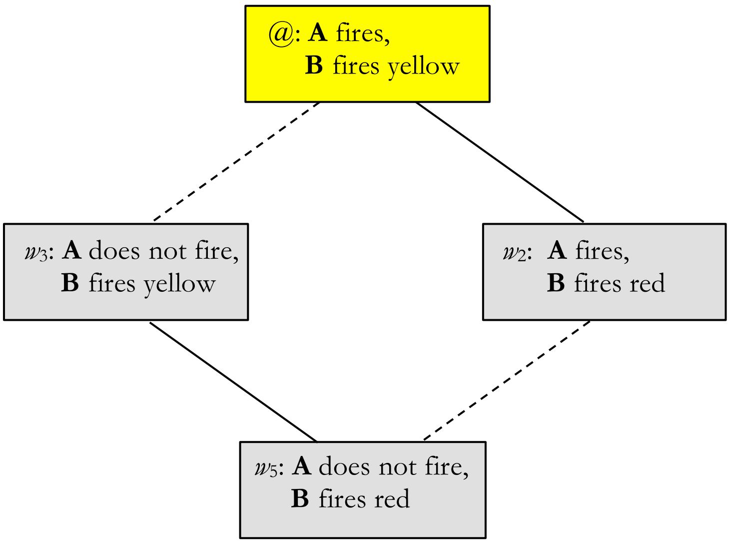
Possibility horizon $\mathscr{H}_4$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *G-yellow* rather than *G-green* ?	$\mathscr{H}_4$	✓	–
Does *B-yellow* cause *G-yellow* rather than *G-green*?	$\mathscr{H}_4$	✓	✓

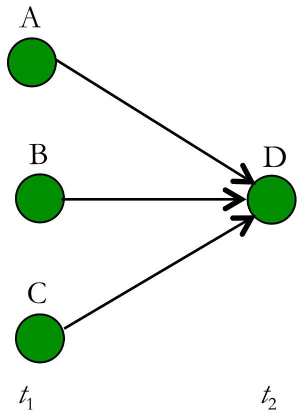
Finally, consider the following quaternary causal claim '*B-yellow* rather than *B-red* caused *G-yellow* rather than ¬*G*'. This requires us to consider the possibility horizon $\mathscr{H}_5$ illustrated below:

Possibility horizon $\mathscr{H}_5$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *B-yellow* rather than *B-red* cause *G-yellow* rather than ¬*G*?	$\mathscr{H}_5$	✓	✓

34.* Overdetermined joint causation[34]

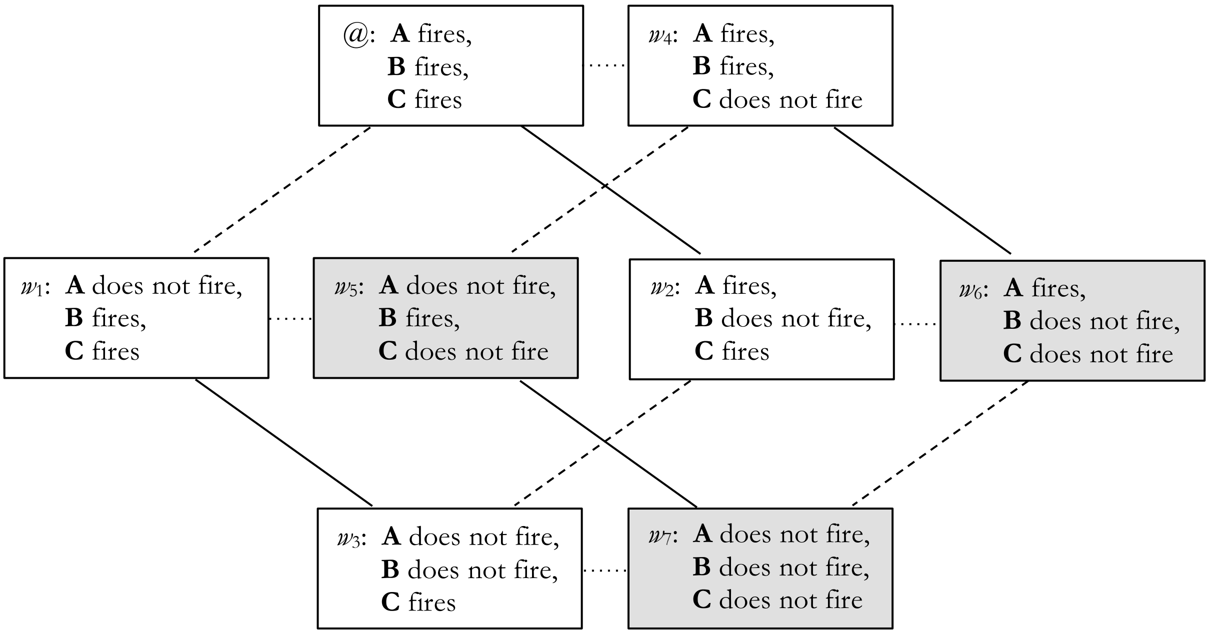


Note: **D** fires if and only if it receives signals from both **A** and **B**, or from **C**.

We here consider the effect *D* and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which *D* does not happen are coloured grey:

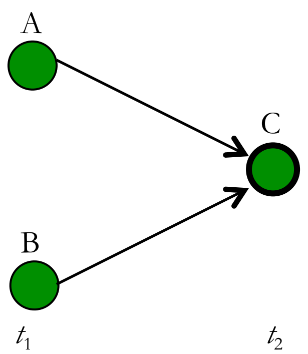
[34] I have constructed this case to illustrate a particular point about the statement of the intrinsicness principle; I have not seen it discussed elsewhere.

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does *A* cause *D*?	$\mathscr{H}$	✓	✓
Does *B* cause *D*?	$\mathscr{H}$	✓	✓
Does *C* cause *D*?	$\mathscr{H}$	✓	✓

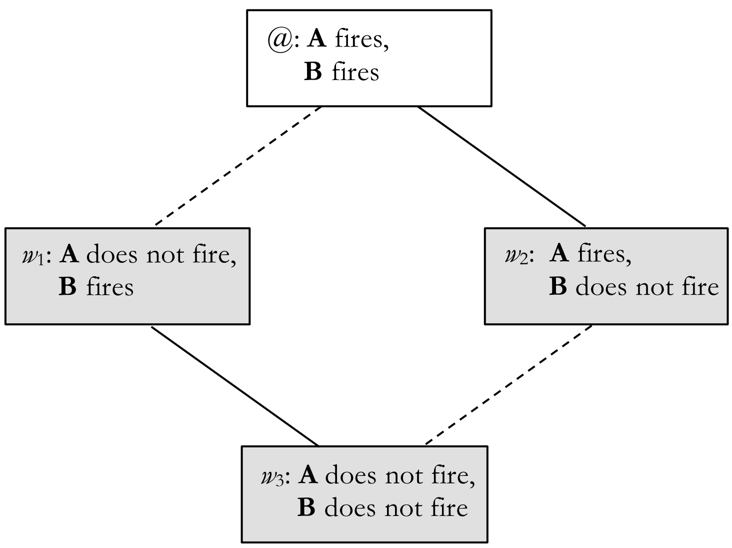
35. Joint causation[35]



Note: **C** is a stubborn neuron that requires two stimulatory signals in order to fire.

We here consider the effect C and its causes at time t_1 within the possibility horizon illustrated below – where the worlds in which C does not happen are coloured grey:

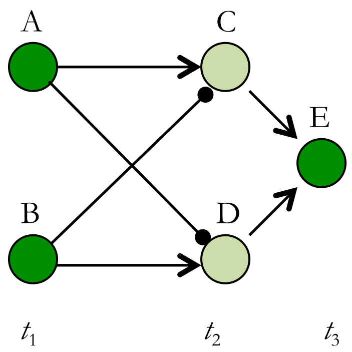
Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause C?	$\mathscr{H}$	✓	✓
Does B cause C?	$\mathscr{H}$	✓	✓

[35] This is a very familiar causal structure, not explicitly illustrated in Paul and Hall (2013).

36. Complex joint causation[36]



In this case, neurons can fire with different intensities. When a neuron is coloured dark green, this indicates that it fires with intensity 2. When a neuron is coloured light green, this indicates that it fires with intensity 1. In the figure, **A** fires with intensity 2, and sends a stimulatory signal of intensity 2 to **C** and an inhibitory signal of intensity 1 to **D**. **B** similarly fires with intensity 2, and sends a stimulatory signal of intensity 2 to **D** and an inhibitory signal of intensity 1 to **C**. If **A** had fired with intensity 1, it would have sent a stimulatory signal of intensity 1 to **C** and no inhibitory signal to **D**, and similarly in the case of **B**.

The rule for calculating the intensity with which **C** fires is this: **C**'s intensity is given by the intensity of **A**'s stimulatory signal *minus* the intensity of **B**'s inhibitory signal. The corresponding rule applies in the case of **D**.

Finally, **E** will fire if and only if it receives stimulatory signals of a combined intensity of at least 2.

Let *A* be the event that is essentially **A**'s firing, and let *B* be the event that is essentially *B*'s firing. Furthermore, let *A-2* be the event that is essentially *A*'s firing with intensity 2, and let *B-2* be the event that is essentially *B*'s firing with intensity 2.

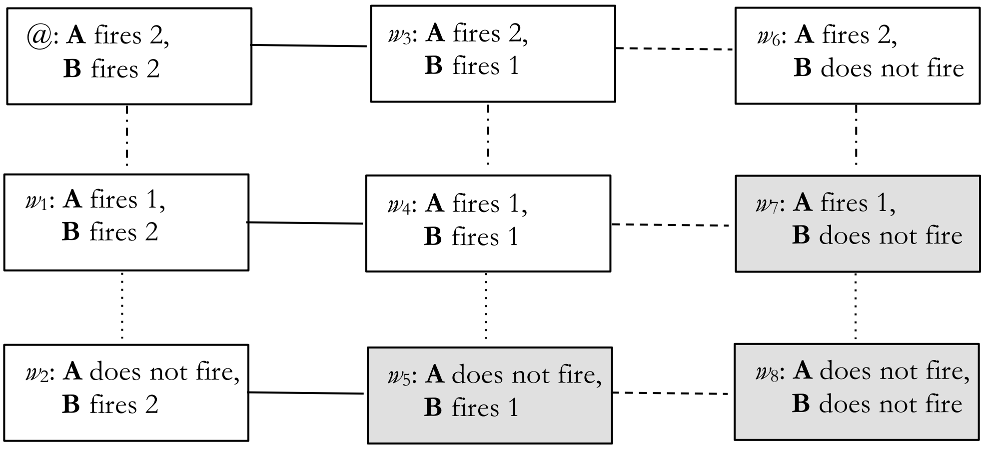
The following is a real-life case with this structure:

> *Moving furniture*: Suzy and Billy move a sofa together. Each of them is strong enough to move it on their own. However, since they are doing it together, they both scale back their efforts.

[36] Cf. Paul and Hall (2013), Figure 16.

We here consider the effect E and its causes at time t_1. The relevant possibility horizon is as illustrated below – where the worlds in which E does not happen are coloured grey:

Possibility horizon $\mathscr{H}$:



Question	Possibility horizon	Process-connection	Security-dependence
Does A cause E?	$\mathscr{H}$	✓	✓
Does A-2 cause E?	$\mathscr{H}$	–	✓
Does B cause E?	$\mathscr{H}$	✓	✓
Does B-2 causes E?	$\mathscr{H}$	–	✓